N.K. Nikolskij (Ed.)

Functional Analysis I

Linear Functional Analysis

Springer-Verlag

Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest
List of Editors, Authors and Translators

Editor-in-Chief

R.V. Gamkrelidze, Academy of Sciences of the USSR, Steklov Mathematical Institute, ul. Vavilova 42, 117966 Moscow, Institute for Scientific Information (VINITI), ul. Usievicha 20a, 125219 Moscow, USSR

Consulting Editor

N.K. Nikol'skij, Steklov Mathematical Institute, Fontanka 27, St. Peterburg D-111, USSR

Author

Yu.I. Lyubich, Department of Mathematics, State University of New York at Stony Brook, Stony Brook, NY 11794-3651, USA

Translator

I. Tweddle, Department of Mathematics, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, Scotland
# Linear Functional Analysis

Yu.I. Lyubich

Translated from the Russian
by I. Tweddle

## Contents

1. Preface .................................................. 4

Chapter 1. Classical Concrete Problems ......................... 6

1. Elementary Analysis ...................................... 6
   1.1. Differentiation .................................... 6
   1.2. Solution of Non-linear Equations .................. 7
   1.3. Extremal Problems .................................. 10
   1.4. Linear Functionals and Operators .................... 10
   1.5. Integration ........................................ 11
   1.6. Differential Equations ............................. 16

2. The Fourier Method and Related Questions ..................... 19
   2.1. The Vibrations of a String ............................ 19
   2.2. Heat Conduction .................................... 21
   2.3. The Classical Theory of Fourier Series .............. 21
   2.4. General Orthogonal Series ........................... 27
   2.5. Orthogonal Polynomials .............................. 28
   2.6. The Power Moment Problem ........................... 29
   2.7. Jacobian Matrices .................................. 34
   2.8. The Trigonometric Moment Problem ................. 35
   2.9. The Fourier Integral ................................ 36
   2.10. The Laplace Transform ............................ 39
   2.11. The Sturm-Liouville Problem ....................... 45
   2.12. The Schrödinger Operator on the Semiaxis .......... 47
   2.13. Almost-Periodic Functions .......................... 51

3. Theory of Approximation .................................. 53
   3.1. Chebyshev Approximations ............................ 53
   3.2. Chebyshev and Markov Systems ...................... 56
   3.3. The Chebyshev-Markov Problem ....................... 58
Contents

3.4. The $L$-problem of Moments ........................................ 60
3.5. Interpolation and Quadrature Processes .......................... 61
3.6. Approximation in the Complex Plane ................................ 65
3.7. Quasianalytic Classes ............................................... 69

§4. Integral Equations .................................................... 70
4.1. Green's Function .................................................... 70
4.2. Fredholm and Volterra Equations ................................. 74
4.3. Fredholm Theory ..................................................... 76
4.4. Hilbert-Schmidt Theory ............................................ 77
4.5. Equations with Difference Kernels ............................... 78
4.6. The Riemann-Hilbert Problem ..................................... 82

Chapter 2. Foundations and Methods ................................... 85

§1. Infinite-Dimensional Linear Algebra ............................... 85
1.1. Bases and Dimension ............................................... 85
1.2. Homomorphisms and Linear Functionals .......................... 88
1.3. The Algebraic Theory of the Index ............................... 93
1.4. Systems of Linear Equations ..................................... 95
1.5. Algebraic Operators ............................................... 97
1.6. General Principles of Summation of Series ..................... 100
1.7. Commutative Algebra .............................................. 101

§2. Convex Analysis ...................................................... 107
2.1. Convex Sets ......................................................... 107
2.2. Convex Functionals ............................................... 109
2.3. Seminorms and Norms ............................................. 111
2.4. The Hahn-Banach Theorem ...................................... 112
2.5. Separating Hyperplanes .......................................... 113
2.6. Non-negative Linear Functionals ................................ 115
2.7. Ordered Linear Spaces ............................................ 117

§3. Linear Topology ...................................................... 119
3.1. Linear Topological Spaces ....................................... 119
3.2. Continuous Linear Functionals ................................... 127
3.3. Complete Systems and Topological Bases ....................... 132
3.4. Extreme Points of Compact Convex Sets ....................... 138
3.5. Integration of Vector-Functions and Measures ................. 140
3.6. $w^*$-Topologies ................................................... 144
3.7. Theory of Duality .................................................. 147
3.8. Continuous Homomorphisms ..................................... 153
3.9. Linearisation of Mappings ....................................... 163

§4. Theory of Operators .................................................. 165
4.1. Compact Operators ................................................ 165
4.2. The Fixed Point Principle ....................................... 174
4.3. Actions and Representations of Semigroups .................... 178
4.4. The Spectrum and Resolvent of a Linear Operator ............ 181
4.5. One-Parameter Semigroups ........................................ 186
4.6. Conjugation and Closure ........................................ 192
4.7. Spectra and Extensions of Symmetric Operators ............... 194
4.8. Spectral Theory of Selfadjoint Operators ...................... 197
4.9. Spectral Operators ................................................ 206
4.10. Spectral Subspaces ............................................... 207
4.11. Eigenvectors of Conservative and Dissipative Operators ... 211
4.13. Complete Compact Operators .................................. 216
4.14. Triangular Decompositions .................................... 220
4.15. Functional Models ............................................... 222
4.16. Indefinite Metric .............................................. 227
4.17. Banach Algebras ................................................ 232
§ 5. Function Spaces .................................................. 242
5.1. Introductory Examples .......................................... 242
5.2. Generalised Functions .......................................... 244
5.3. Families of Function Spaces .................................... 252
5.4. Operators on Function Spaces .................................. 255

Commentary on the Bibliography .................................... 261

Bibliography ........................................................ 261

Author Index ....................................................... 273

Subject Index ..................................................... 277
Preface

Up to a certain time the attention of mathematicians was concentrated on the study of individual objects, for example, specific elementary functions or curves defined by special equations. With the creation of the method of Fourier series, which allowed mathematicians to work with arbitrary functions, the individual approach was replaced by the “mass” approach, in which a particular function is considered only as an element of some function space. More or less simultaneously the development of geometry and algebra led to the general concept of a linear space, while in analysis the basic forms of convergence for series of functions were identified: uniform, mean square, pointwise and so on. It turns out, moreover, that a specific type of convergence is associated with each linear function space, for example, uniform convergence in the case of the space of continuous functions on a closed interval. It was only comparatively recently that in this connection the general idea of a linear topological space (LTS)¹ was formed; here the algebraic structure is compatible with the topological structure in the sense that the basic operations (addition and multiplication by a scalar) are continuous. Included in this scheme are spaces which, historically, had appeared earlier, namely Fréchet spaces (metric with a complete translation invariant metric), Banach spaces (complete normable) and finally the class which is the most special of all but at the same time the most important for applications, Hilbert spaces, whose topology and geometry are defined in a manner which goes back essentially to Euclid – the assignment of a scalar product of vectors.

Using contemporary formal language we can say that LTSs form a category in which continuous homomorphisms, or continuous linear operators (we usually apply the last term to homomorphisms of a space into itself), serve as morphisms. Specific classical examples of linear operators are differentiation and integration or, in a more general form, differential and integral operators. As a rule, integral operators are continuous but this cannot be said of differential operators. Thanks to the construction of a sufficiently general theory of linear operators it became possible to include the latter case.

The story was repeated at the operator level. At first mathematicians studied individual operators but later on it turned out to be useful and necessary to pass to classes. First and foremost in this connection, multiplication of operators (as a rule non-commutative) went out and algebras of operators appeared. Moreover many natural function spaces are also algebras (commutative, of course: typical multiplication is the usual one, i.e. pointwise, or its Fourier equivalent – convolution on an Abelian group). With regard to topology all these situations are covered by the concept of a topological algebra but with a considerable excess of generality. A satisfactory approach is achieved in the narrower setting of

¹ Translator's note. I will use LTSs for the plural and write an LTS rather than the correct a LTS for the indefinite form since the former reads more smoothly. Other abbreviations of similar type will be treated in the same way.
Banach algebras which have proved to be extremely fruitful in harmonic analysis, in representation theory, in approximation theory and so on.

The development of functional analysis ran its course under the powerful influence of theoretical and mathematical physics. Here we may mention, for example, spectral theory, which evolved from wave mechanics, the technique of generalised functions (or distributions), whose construction and widespread introduction was preceded by the systematic practice of using the $\delta$-function in quantum mechanics, ergodic theory, whose fundamental problems were posed by statistical physics, the investigation of operator algebras in connection with applications to quantum field theory and statistical physics and so on. At the present time the ideas, terminology and methods of functional analysis have penetrated deeply not only into natural science but also into such applied disciplines as numerical mathematics and mathematical economics.

In the introductory volume presented below the classical sources of functional analysis are traced, its basic core is described (with a sufficient degree of generality but at the same time with a series of concrete examples and applications) and its principal branches are outlined. An expanded account of a series of specific sections will be given in the subsequent volumes, while certain questions closely connected with linear functional analysis have already been elucidated in previous volumes of the present series. We only touch fragmentarily upon non-linear aspects.

The general plan of the volume was discussed with R.V. Gamkrelidze and N.K. Nikol'skij and individual topics with A.M. Vershik, E.A. Gorin, M.Yu. Lyubich, A.S. Markus, L.A. Pastur and V.A. Tkachenko. Valuable information on certain questions which are elucidated in the volume was kindly provided to the author by V.M. Borok, Yu.A. Brudnyj, V.M. Kadets, M.I. Kadets, V.Eh. Katsnelson and Yu.I. Lyubarskij. The author offers profound thanks to all the named individuals. He is also most grateful to Dr. Ian Tweddle for his translation of the work into English.
Chapter 1
Classical Concrete Problems

§ 1. Elementary Analysis

The overwhelming majority of functional relations encountered in mathematics and its applications are non-linear 'in the large' but linear 'in the small'. For example the linear theory of elasticity is founded on the assumption that stresses depend linearly on deformations when the latter are small. The method of mathematical analysis, which is firmly established in the works of Newton and Leibniz, consists of two basic procedures:

local linearisation of the functions being studied (differentiation) and reconstruction of the global relation according to its local structure (integration).

Functional analysis was conceived in the womb of classical mathematical analysis but it took shape as an independent discipline only at the end of the 19th and the beginning of the 20th centuries.

1.1. Differentiation. If we speak about classical analysis using modern language, we can say that one of its fundamental objects is a differentiable real or complex function of \( n \) real variables, i.e. a mapping \( f: G \to \mathbb{C} \) (\( G \) a domain in Euclidean space \( \mathbb{R}^n \)), having the property that at each point \( x \in G \) the increment \( f(x + h) - f(x) \) as a function of the displacement vector \( h \) can be approximated by a linear function to within order \( o(|h|) \) as \( |h| \to 0 \):

\[
f(x + h) - f(x) = (g(x), h) + o(|h|). \tag{1}
\]

The uniquely determined vector \( g(x) \) in this situation is the gradient \( \nabla f(x) \), whose coordinates are the partial derivatives \( \frac{\partial f}{\partial x_k} \) (\( 1 \leq k \leq n \)). The principal linear part of (1) is by definition the differential \( df \) of the function \( f \) at the point \( x \). In classical notation \( h = dx \) and

\[
df = (\nabla f(x), dx) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} dx_k.
\]

Thus the function \( f \) is locally linearised:

\[
f(x + dx) \approx f(x) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} dx_k.
\]

The function \( f \) is said to be smooth if its partial derivatives exist everywhere in \( G \) and are continuous. Any smooth function is differentiable. If a function has continuous partial derivatives of all forms up to order \( r \) inclusive, then we say

---

1 Not only direct integration but also indirect processes such as the solution of differential equations.
that it belongs to the class\(^2\) \(C'(G)\). The class \(C^\infty(G)\) is defined as the intersection of all the \(C'(G)\). The enumerated classes can also be considered on closed domains, for example on the interval \([a, b] \subset \mathbb{R}\) (and even on arbitrary subsets \(X \subset \mathbb{R}^n\)).

If the function \(f\) belongs to \(C'(G)\) \((r < \infty)\) then at any point \(x \in G\) we have Taylor's formula

\[
f(x + dx) = f(x) + \sum_{m=1}^{r} \frac{1}{m!} \sum_{i_1 + \ldots + i_n = m} \frac{\partial^m f}{\partial x_{i_1} \ldots \partial x_{i_n}} \, dx_{i_1} \ldots dx_{i_n} + o(|dx|^r),
\]

i.e. to within order \(o(|dx|^r)\) the function \(f\) is a polynomial of \(r\)th degree in \(dx\).

The inner sum in Taylor's formula is the \(m\)th differential of the function \(f\) at the point \(x\).

For a function \(f \in C^\infty(G)\) we can formally write down its Taylor series by passing in Taylor's formula to \(r = \infty\) (and of course rejecting the error term). If the Taylor series of a function \(f \in C^\infty(G)\) converges to it in some neighbourhood of the point \(x \in G\), then \(f\) is said to be analytic at the point \(x\). A function which is analytic at all points of the domain \(G\) is said to be analytic on \(G\). A function which is analytic on the whole space is said to be an entire function.

All that has been said above carries over immediately to complex functions of \(n\) complex variables. In the 18th century the latter situation had already begun to separate out little by little into a distinct branch of mathematical analysis (complex analysis) in view of the fact that complex analytic functions of one or several complex variables have very distinctive properties. In particular, from the definition of an analytic function (complex or real) one easily obtains the uniqueness theorem: an analytic function which is flat at some point (i.e. such that all its derivatives are equal to zero at this point) is identically zero.

The standard example of an infinitely differentiable non-zero function which is flat (at \(x = 0\)) is

\[
f(x) = \begin{cases} \frac{x^2}{x^2 + 1} & (x \neq 0), \\ 0 & (x = 0). \end{cases}
\]

In what follows, either it will be stated or it will be clear from the context whether the functions under consideration and their independent variables are real or complex.

1.2. Solution of Non-linear Equations. This is one of the first applications of linearisation. Let us consider the equation \(f(x) = 0\), where \(f\) is a real function which is smooth on a certain interval \([a, b] \subset \mathbb{R}\). The existence of a root is guaranteed if, for example, \(f(a) < 0\) and \(f(b) > 0\). The uniqueness of the root is guaranteed if the derivative \(f'(x) > 0\) for all \(x\). Assuming these conditions are satisfied, let us denote the root by \(\bar{x}\). If some approximation \(x_0\) to the root \(\bar{x}\) is known then, using (1), we obtain

\[
f(x_0) + f'(x_0)(\bar{x} - x_0) \approx 0,
\]

\(^2\)For convenience we also put \(C^0(G) \equiv C(G)\), where \(C(G)\) is the class of continuous functions on \(G\).
Chapter 1. Classical Concrete Problems

from which

\[ \bar{x} \approx x_1 \equiv x_0 - \frac{f(x_0)}{f'(x_0)}. \]  

(2)

In exactly the same way the next approximation \( x_2 \) is formed from \( x_1 \) (if \( a < x_1 < b \)) and so on. This iterative process is called Newton's process or the method of tangents, since formula (2) corresponds to approximating the graph of the function \( f \) by the tangent to it at the point \((x_0, f(x_0))\).

**Theorem 1.** Let \( f \) be a function of class \( C^2[a, b] \) satisfying the conditions:

\[ f(a) < 0, \quad f(b) > 0; \quad f'(x) > 0, \quad |f''(x)| \leq M. \]

If

\[ |x_0 - \bar{x}| \leq \frac{2\mu}{5M} q, \]  

(3)

where \( q < 1 \), then for the initial approximation \( x_0 \), Newton's process is defined on the interval \([a, b]\) and converges to the root \( x \) with bound

\[ |x_m - \bar{x}| \leq \frac{2\mu}{5M} q^{2m} \quad (m = 0, 1, 2, \ldots). \]  

(4)

Thus the linearisation of the original equation generates an iterative process which converges very rapidly\(^3\) to the root. We note however that the initial approximation \( x_0 \) must be chosen with sufficient accuracy according to (3). Condition (3) cannot be completely effective since the unknown root \( \bar{x} \) appears in it. This deficiency can be easily removed but the local character of Newton's process is essentially connected with its construction by means of linearisation.

**Example.** Let \( f(x) = x^3 + ax + b \) \((-\infty < x < \infty)\). If \( 2b^2 + a^3 = 0 \) and \( a \neq 0 \) then the points \( x_0 = 0, x_1 = -b/a \) form a cycle of the second order for Newton's process corresponding to the function \( f \). For any initial approximation close to \( x_0 \) or \( x_1 \), the process converges on the stated cycle and not to the root of the equation \( f(x) = 0 \). This continues to be the case under small perturbations of the initially chosen values of the parameters \( a, b \).

Any mapping \( f: \mathbb{R} \rightarrow \mathbb{R} \) generates an iterative process

\[ x_{m+1} = f(x_m) \quad (m = 0, 1, 2, \ldots). \]

If we only know that \( f \) is continuous then the behaviour of the iterations as \( m \to \infty \) can turn out to be extremely complicated. However under certain stronger assumptions about \( f \) global convergence of the process can take place.

\(^3\)The rate of convergence defined by the bound (4), i.e. \( O(q^m) \), is said to be quadratic while with bound \( O(q^m) \) it is linear; in such cases we may call the process rapidly convergent in contrast to slowly convergent processes for which a typical bound is \( O(m^{-\alpha}) \) \((\alpha > 0)\).
Theorem 2. If the function $f$ satisfies the Lipschitz condition
\[ |f(x_1) - f(x_2)| \leq q|x_1 - x_2| \] (5)
with $0 < q < 1$, then the iterative process defined by it converges for any initial approximation with bound
\[ |x_m - \bar{x}| \leq |x_0 - \bar{x}|q^m, \]
where $\bar{x}$ is the root of the equation $x = f(x)$. The existence and uniqueness of the root is guaranteed by condition (5).

A sufficient condition for (5) to hold is that $f$ should be differentiable and $|f'(\bar{x})| \leq q$.

Theorem 2 is widely used in applications.

Example. In celestial mechanics we encounter Kepler's equation
\[ x - \varepsilon \sin x = \zeta \] (6)
with unknown $x$ and parameters $\varepsilon, \zeta$ ($0 \leq \varepsilon < 1; -\infty < \zeta < \infty$). Since the function $f(x) = \varepsilon \sin x + \zeta$ satisfies the conditions of the previous theorem, equation (6) has a unique real root $\bar{x}$, which can be found by means of the iterations
\[ x_{m+1} = \varepsilon \sin x_m + \zeta \quad (m = 0, 1, 2, \ldots) \]
with any $x_0$.

Theorem 2 extends without change in statement (or proof) to mappings of the semiaxis or of any interval into itself4. This is important because many of its applications are of a local character.

Example (Implicit function theorem). Let $f(x_1, x_2)$ be a function of class $C^2$ in a neighbourhood of the point $(0, 0) \in \mathbb{R}^2$ such that $f(0, 0) = 0$. Then if $\frac{\partial f}{\partial x_2}(0, 0) \neq 0$, there exists on some neighbourhood $|x_1| < \eta$ a unique function $x_2 = \phi(x_1)$ which satisfies the equation $f(x_1, x_2) = 0$ and the condition $\phi(0) = 0$.

In fact the equation under consideration can be written in the form
\[ x_2 = \lambda x_1 + \rho(x_1, x_2), \] (7)
where $\lambda$ is a constant, $\rho|_{(0, 0)} = 0$ and $\nabla \rho|_{(0, 0)} = 0$. We seek the root $x_2$ by considering $x_1$ as a parameter belonging to a neighbourhood of zero. Since the derivative with respect to $x_2$ is in modulus less than any given $q$ ($0 < q < 1$) on a sufficiently small neighbourhood of the point $(0, 0)$, we may choose $\eta > 0$ and $M > 0$ such that for $|x_1| < \eta$, the interval $|x_2| \leq M\eta$ is mapped into itself by the function (7). By Theorem 2 equation (7) has a unique root $x_2 = \phi(x_1)$ for $|x_1| < \eta$.

---

4In general, if $X$ is a complete metric space and the mapping $f: X \to X$ contracts uniformly, i.e. $\text{dist}(fx, fy) \leq q \text{dist}(x, y)$ where $q < 1$, then the iterative process $x_{m+1} = fx_m$ ($m = 0, 1, 2, \ldots$) converges for any $x_0$ to the unique root $\bar{x}$ of the equation $x = fx$ i.e. to the fixed point of the mapping $f$; moreover we have the bound $\text{dist}(x_m, \bar{x}) \leq q^m \text{dist}(x_0, \bar{x})$ ($m = 0, 1, 2, \ldots$) (contraction mapping principle).
1.3. Extremal Problems. Let the function $f: G \to \mathbb{R}$ ($G$ a domain in $\mathbb{R}^n$) be differentiable at the point $x \in G$. If it attains a local extremum at this point then $df \equiv 0$ which is equivalent to the equation

$$\nabla f(x) = 0,$$

i.e. $x$ is a stationary (critical) point for $f$. This necessary condition for an extremum is one of the first basic results of differential calculus. Its discovery is due to Fermat (1643). Further development in this direction came about on the creation of variational calculus (Johann Bernoulli, Euler, Lagrange).

In variational calculus we consider functions of variables which are of a more complicated nature than points in $\mathbb{R}^n$. For example, in the classical brachistochrone problem it is required to find the curve joining two given points in a vertical plane such that a heavy particle moving along it takes the least possible time. Here the time of motion is a function on the set of smooth curves joining the given points. Functions of such a type are given the name functionals. More abstractly a functional is defined as a real or complex function on any set $X$ (i.e. as a mapping of $X \to \mathbb{R}$ or $X \to \mathbb{C}$).

In variational calculus necessary conditions of the type (8) are established for a wide range of problems on the extremum of functionals. Condition (8) in coordinates represents a system of $n$ equations with $n$ unknowns $x_1, \ldots, x_n$:

$$\frac{\partial f}{\partial x_k} = 0 \quad (1 \leq k \leq n).$$

The corresponding conditions in variational calculus are ordinary or partial differential equations. They can have deep independent meaning. For example, the dynamics of a wide class of mechanical systems are described by equations of this type; they arise in electrodynamics, in the theory of relativity and so on. In physical theory constructed in the standard way we encounter a certain functional called action and the fundamental equations of the theory can be written as the necessary conditions for an extremum of the action. This approach is called the variational principle of the given theory (or the principle of least action, although the attainment of even a local extremum on trajectories is not at all necessary and therefore it is better to speak of the principle of stationary action).

1.4. Linear Functionals and Operators. The abstract interpretation of functions as elements (points) of some set (usually a real or complex linear space) is fundamental to functional analysis. Such an approach, whose pioneers were Volterra and Hadamard, opened up exceptionally wide possibilities for applying geometric and algebraic methods in analysis.

The following spaces may serve as initial examples of linear function spaces (real or complex): the spaces $C^r(G)$ ($0 \leq r \leq \infty; G \subset \mathbb{R}^n$), the space $B(X)$ of bounded functions on an arbitrary set $X$, the space $C(X)$ of continuous functions on a topological space $X$, the space $CB(X) = C(X) \cap B(X)$. They are all infinite-
dimensional as in general are the great majority of function spaces\(^5\) (with the exception of the case where functions on a finite set are being considered).

The simplest class of functionals on a linear space is the class of linear functionals, i.e. additive, homogeneous functionals:

\[ \psi(f_1 + f_2) = \psi(f_1) + \psi(f_2), \quad \psi(af) = a\psi(f) \quad \text{for any scalar } a. \]

Using standard algebraic terminology, we can say that a linear functional is a homomorphism of the linear space into its ground field. A homomorphism \(A\) of a linear space \(E\) into itself (or, in a more general form, a homomorphism of a subspace \(L \subset E\) into \(E\)) is said to be a linear operator on \(E\) (with domain of definition \(D(A) = L\)).

**Example 1.** Let us consider the space \(C[a, b]\) of continuous functions on the interval \([a, b]\). Every point \(x_0 \in [a, b]\) defines a linear functional \(\delta_{x_0}[f] = f(x_0)\). This functional (with \(x_0 = 0\)) is called the \(\delta\)-function in the terminology of Dirac who systematically worked with it in quantum mechanics\(^6\).

**Example 2.** In the space \(C[a, b]\) multiplication of a function by any fixed continuous function \(g(x)\) (in particular, by \(x\)) is a linear operator defined on the whole space.

**Example 3.** In the space \(C^\infty(G)\) any vector field \((a_1(x), \ldots, a_n(x)) (x \in G)\) with components \(a_k(x) \in C^\infty(G)\) defines a linear operator

\[ (Df)(x) = \sum_{k=1}^{n} a_k(x) \frac{\partial f}{\partial x_k}. \]

**Example 4.** In the space \(C[a, b]\) differentiation \(f \mapsto \frac{df}{dx}\) is a linear operator with domain of definition \(C^1[a, b]\). If desired we can restrict the domain of definition by imposing a boundary condition, e.g. \(f(a) = 0\) or \(f(a) = f(b)\).

Clearly this procedure changes the operator although its effect on a function that remains in its domain of definition is not altered. This way of considering supplementary conditions is essential for the application of the abstract theory of operators.

The theory of linear operators is one of the fundamental divisions of functional analysis.

1.5. **Integration.** The classical figures of mathematical analysis understood the definite integral as the sum of an infinitely large number of infinitely small

---

\(^5\) A function space can be formally defined as any subspace of the space \(\Phi(X)\) of all (real or complex) functions on an arbitrary set \(X\). In particular, if \(X = \mathbb{Z}\) (the set of all integers) or \(X = \mathbb{N} = \{0, 1, 2, \ldots\}\), then \(\Phi(X)\) becomes the space of sequences (two-sided or one-sided in the respective cases).

\(^6\) Dirac defined the \(\delta\)-function by means of the formal requirements \(\delta(x) = 0\) (\(x \neq 0\)), \(\delta(0) = \infty\), \(\int_{-\infty}^{\infty} \delta(x) \, dx = 1\), which give rise to the integral representation of the identity operator: \(f(x) = \int_{-\infty}^{\infty} \delta(x - y)f(y) \, dy\); the \(\delta\)-function is the prototype of the modern concept of generalised function (or distribution).
Chapter 1. Classical Concrete Problems

elements. Cauchy (1823) gave this idea a precise meaning when applied to continuous functions. Then Riemann (1854) introduced the class of all functions for which Cauchy's definition of the integral is meaningful. For continuous functions Stieltjes (1894) introduced integration with respect to a weight function which allowed the connection to be made between integration and discrete summation. These achievements of the 19th century were superseded in the 20th century by the works of Lebesgue and his successors, as a result of which the general theory of measure and integration was created. This was accompanied by an important extension of the collection of function spaces: there appeared the space $L^1(X, \mu)$ of Lebesgue integrable functions on any set (space) $X$ endowed with some measure $\mu$, the space $L^2(X, \mu)$ of measurable functions the square of whose modulus is in $L^1(X, \mu)$ and, following this, the entire family $L^p(X, \mu)$ ($1 \leq p < \infty$) of spaces of measurable functions for which the $p$th power of the modulus belongs to $L^1(X, \mu)$.

A measure $\mu$ on $X$ is a countably additive function on some $\sigma$-algebra of subsets of the set $X$ which takes values in $[0, \infty]$. If $\mu(X) < \infty$ the measure is said to be finite. The sets appearing in the associated $\sigma$-algebra are said to be measurable. A real function $f$ on $X$ is said to be measurable if for any half open interval $[a, b) \subset \mathbb{R}$ its inverse image $f^{-1}[a, b)$ is a measurable subset of $X$.

Given a measurable function $f \geq 0$, i.e. $f(x) \geq 0$ for all $x \in X$, then its Lebesgue integral with respect to the measure $\mu$ is defined as

$$\int_X f \, d\mu = \sup \sum_{k=0}^{\infty} a_k \mu(f^{-1}[a_k, a_{k+1}]).$$

where the supremum is taken over all sequences $0 = a_0 < a_1 < a_2 < \ldots, a_k \to \infty$. This integral is a non-negative number or $+\infty$. In the first case the function $f$ is Lebesgue integrable or summable.

For a real measurable function whose sign is unrestricted we put $f_+ = \max(f, 0)$, $f_- = -\min(f, 0)$ and say that $f$ is summable if both functions $f_+, f_-$ are summable, or equivalently $|f|$ is summable. Further, by definition,

$$\int_X f \, d\mu = \int_X f_+ \, d\mu - \int_X f_- \, d\mu.$$

Finally, if $f$ is complex, $f = f_1 + if_2$, then $f$ is said to be summable if $f_1$ and $f_2$ are summable and also by definition

---

7 In modern times the classical concept has been revived with non-standard analysis.

8 A $\sigma$-algebra is any family of subsets containing $X$ and closed with respect to the operations of complementation and countable union.

9 By definition measurability of a complex function reduces to measurability of its real and imaginary parts.

10 The basic concepts of the theory of probability may be introduced on the basis of measure theory: a probability space is a set $X$ with a normalised measure $\mu (\mu(X) = 1)$, called a probability distribution (or probability measure) on $X$; random events are the measurable sets $A \subset X$ ($\mu(A)$ is the probability of the event $A$); random variables are measurable functions and so on. This approach, which is now generally accepted, was proposed by A.N. Kolmogorov in the early 1930s.
Along with each function \( g \in L^1(X, \mu) \) every measurable function \( f \) such that \(|f| \leq |g|\) is summable. It therefore follows from the elementary inequality
\[
|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p)
\]
that all the spaces \( L^p(X, \mu) \) \((p \geq 1)\) are linear.

If \( f \in L^1(X, \mu) \) and
\[
\int_X |f| \, d\mu = 0
\]
then \( f(x) \neq 0 \) only on a set of measure zero or, as it is usually expressed, \( f(x) = 0 \) almost everywhere. Equality almost everywhere is an equivalence relation on the set of all measurable functions; moreover equivalent functions \( f_1, f_2 \) are simultaneously summable or not summable and in the first case
\[
\int_X f_1 \, d\mu = \int_X f_2 \, d\mu.
\]
This allows us to identify them in what follows. From now on \( L^1(X, \mu) \) and along with it all the \( L^p(X, \mu) \) are in fact converted into linear spaces which consist of the classes under the stated equivalence relation.

The space \( L^2(X, \mu) \) occupies a special place in the family \( L^p(X, \mu) \) \((p \geq 1)\) because of the fact that on it there is a natural scalar product
\[
(f_1, f_2) = \int_X f_1 \bar{f_2} \, d\mu
\]
(this definition is meaningful by virtue of the inequality \(|ab| \leq \frac{1}{2}(|a|^2 + |b|^2)|\)). All the properties required in linear algebra of a scalar product are fulfilled here. Therefore in \( L^2(X, \mu) \) we have a Euclidean geometry (in general, it is infinite-dimensional). In particular we can speak of orthogonal functions \( f_1, f_2: (f_1, f_2) = 0 \). A function \( f \) is said to be normalised if \((f, f) = 1\). A system of functions \( \{f_i\}_1^\infty \) is said to be orthogonal if \((f_i, f_k) = 0\) \((i \neq k)\) and orthonormal if \((f_i, f_k) = \delta_{ik}\).

Pythagoras Theorem. If \( f_1, f_2 \) are orthogonal then
\[ \|f_1 + f_2\|^2 = \|f_1\|^2 + \|f_2\|^2. \]

In any \( L^p(X, \mu) \) \((p \geq 1)\) there is also a natural norm
\[ \|f\| = \left( \int_X |f|^p \, d\mu \right)^{1/p}, \]
but for \( p \neq 2 \) it is not connected with any scalar product. The corresponding triangle inequality is called Minkowski's inequality. As a result for any \( p \geq 1 \) we may turn \( L^p(X, \mu) \) into a metric space by putting \( d(f_1, f_2) = \|f_1 - f_2\| \).

**Riesz-Fischer Theorem.** All the metric spaces \( L^p(X, \mu) \) are complete.

In any real or complex linear space \( E \) a functional \( \|f\| \) with the following properties is called a norm:
1) \( \|f\| > 0 \) \((f \neq 0)\);
2) \( \|\alpha f\| = |\alpha| \|f\| \) for every scalar \( \alpha \);
3) \( \|f_1 + f_2\| \leq \|f_1\| + \|f_2\| \).

A norm generates a metric (just as in the examples given above) and, consequently, a topology. A linear space provided with a norm is said to be normed. A normed linear space which is complete with respect to the corresponding metric is called a Banach space. Banach not only initiated the use of this important class of spaces but he also developed its theory deeply.\(^\text{11}\)

A Banach space in which the norm is generated by a scalar product (i.e. \( \|f\| = \sqrt{\langle f, f \rangle} \)) is called a Hilbert space\(^\text{12}\) (without the completeness requirement it is called a pre-Hilbert space). The space \( L^2(X, \mu) \) is a Hilbert space and all the spaces \( L^p(X, \mu) \) \((p \geq 1)\) are Banach spaces. The most interesting and furthest developed aspects of functional analysis are connected with Hilbert spaces.

Let us return to the question of integration of a function. Let \( I \) be any interval on the real axis and \( \lambda(x) \) \((x \in I)\) a non-decreasing function. For any bounded interval \((\alpha, \beta) \subseteq I\) we shall call the increment \( \lambda(\beta) - \lambda(\alpha) \) its \( \lambda \)-measure; the jumps of the function \( \lambda \) at the boundary points of the interval \( I \) will be called the \( \lambda \)-measures of these points. By means of this a certain measure \( \mu_\lambda \) can be uniquely defined first of all on the \( \sigma \)-algebra of Borel sets\(^\text{13}\) and then also on the \( \sigma \)-algebra of sets of the form \( B \cup N \) where \( B \) is a Borel set and \( \mu_\lambda^\star(N) = 0 \) \((\mu^\star \) is the outer measure defined by a measure \( \mu \), i.e. \( \mu^\star(A) \) is the infimum of the measures of all measurable sets which contain \( A \)). The latter is called the Lebesgue \( \sigma \)-algebra for a given function \( \lambda \) and the measure \( \mu_\lambda \) is called the Lebesgue-Stieltjes measure defined by the function \( \lambda \). The integral of a function \( f \in L^1(I, \mu_\lambda) \) is usually written in the form

\(^{11}\) Even finite-dimensional normed spaces (all of which are complete) are objects which are highly interesting and also useful for application (for example, in numerical analysis).

\(^{12}\) In honour of Hilbert who introduced (in the setting of \( L^2 \)) methods of Euclidean geometry into analysis.

\(^{13}\) This is the smallest \( \sigma \)-algebra containing the compact sets \( Q \subseteq I \). If a measure on a locally compact topological space \( X \) is defined on all the Borel sets and is finite on all the compact sets then it is called a Borel measure.
If $f$ is continuous then this integral is just the usual Stieltjes integral. Consequently for any $f \in L^1(I, \mu_\lambda)$ the integral (9) is called its Lebesgue-Stieltjes integral. The space $L^p(I, \mu_\lambda)$ is often denoted by $L^p_\lambda(I)$ or more concisely $L^p_\lambda$ (provided the domain of definition appears in the description of the function $\lambda$). In the case $\lambda(x) = x$ the measure $\mu_\lambda$ is called Lebesgue measure. For this measure we write briefly $L^p(I)$ or even simply $L^p$ if the interval $I$ is already fixed.

The theory of Lebesgue also provides a suitable context for the indefinite integral, characterising functions of the form

$$F(x) = \int_a^x f(t) \, dt + \text{const.} \quad (x \in I, \ a \in I)$$

($f$ is locally summable, i.e. $f \in L^1$ on some neighbourhood of each point) by means of the property of absolute continuity: the image of any set of measure zero has measure zero. In this situation $F'(x) = f(x)$ almost everywhere. Any non-decreasing function $\lambda$ is differentiable almost everywhere and $\lambda'$ is locally summable, which allows us to extract from $\lambda$ its absolutely continuous part, the integral of $\lambda'$.

We can introduce analogously the Lebesgue measure and integral on the circle (however this reduces to the case of the interval $[-\pi, \pi]$).

For any two measures $\mu_1, \mu_2$ on spaces $X_1, X_2$ we form naturally their direct product $\mu_1 \times \mu_2$ on $X_1 \times X_2$. For this, if $f \in L^1(X_1 \times X_2, \mu_1 \times \mu_2)$, we have

$$\int_{X_1 \times X_2} f(x_1, x_2) \, d\mu_1 \, d\mu_2 = \int_{X_1} d\mu_1 \int_{X_2} f(x_1, x_2) \, d\mu_2$$

(summability at each step of the repeated integration is guaranteed). This is the theorem of Fubini.

In particular, the construction of Lebesgue measure on $\mathbb{R}^n$ starting from Lebesgue measure on $\mathbb{R}$ is covered by what has been said.

A basic property of the integral on any space $X$ with measure $\mu$ is that the functional

$$\psi(f) = \int_X f \, d\mu$$

on $L^1(X, \mu)$ is linear. In general, integration is a powerful tool for the construction of functionals, both linear and non-linear. For example, the simplest form of functionals arising in variational problems is

$$I[y] = \int_a^b \mathcal{F}(x, y, y') \, dx \quad (y \in C[a, b]),$$

(10)

14 There is in addition the important property of positivity, namely $\psi(f) \geq 0$ if $f \geq 0$. For a linear functional positivity is equivalent to monotonicity: if $f_1 \leq f_2$ then $\psi(f_1) \leq \psi(f_2)$.

15 Also operators.
where \( f \) is a given, sufficiently smooth (not lower than \( C^2 \)) function of three variables. The functional \( I[q] \) for a conservative mechanical system with \( s \) degrees of freedom has form

\[
I[q] = \int_{t_1}^{t_2} I(q, \dot{q}) \, dt,
\]

where \( t \) is time, \( q = q(t) \) \((t_1 \leq t \leq t_2)\) runs through the set of curves of class \( C^2 \) in \( s \)-dimensional configuration space, \( q(t) = (q_1(t), \ldots, q_s(t)) \) and \( L \) is the Lagrange function, i.e. \( L = T - V \) where \( T \) is kinetic energy and \( V \) is potential energy. Action has an analogous structure in other physical theories where it appears as the integral of a certain 'Lagrangian'.

In many situations the integral representation of a functional is an inevitable consequence of its intrinsic properties. For example as F. Riesz showed (1911) any positive linear functional on \( C[a, b] \) has form

\[
\psi(f) = \int_a^b f(x) \, d\lambda(x),
\]

where \( \lambda \) is some non-decreasing function on \([a, b] \). It is essentially uniquely defined, i.e. to within an additive constant on the set of points of continuity. In other words, the measure \( \mu_\lambda \) is uniquely defined.

**Example.** If \( \lambda(x) = 0(x < x_0), \lambda(x) = 1(x > x_0) \) then \( \psi(f) = f(x_0) = \delta_{x_0}[f] \).

1.6. **Differential Equations.** As we have already said above, extremal problems in function spaces are one of the sources of differential equations. Let us consider, for example, the simplest problem on the extremum of a functional of the form (10). The differential (variation) of this functional is equal to

\[
\delta I = \int_a^b \left( f_x \delta y + f_y \delta y' \right) \, dx,
\]

where \( \delta y \) is the increment (variation) of the element \( y \in C^2[a, b] \), \( \delta y' \) is the corresponding variation of the derivative \( y' \), \( f_x \equiv \frac{\partial f}{\partial y} \) and \( f_y \equiv \frac{\partial f}{\partial y'} \). Since \( \delta y' = (\delta y)' \), then, with the assumption that \( y \in C^2[a, b] \), we obtain on integrating by parts,

\[
\delta I = \int_a^b \left( f_y - \frac{d}{dx} f_y' \right) (\delta y) \, dx + f_y' \delta y |_a^b.
\]

For an extremum it is necessary that \( \delta I \equiv 0 \). Since \( \delta y \) is any function of class \( C^2 \), there follows from (13) the Euler-Lagrange equation:

\[
f_y - \frac{d}{dx} f_y' = 0.
\]

This is a second order differential equation with respect to the extremal function \( y \) (under the condition of regularity \( f_{yy'} \neq 0 \)).
Analogously from the variational principle and the expression (11) for action in mechanics we can deduce the equations of motion in generalised coordinates, which are called the Lagrange equations:

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = 0 \quad (1 \leq k \leq s).
\]

Here \( L = T - V \) where

\[
T = \frac{1}{2} \sum_{i,j=1}^{s} \mu_{ij}(q) \dot{q}_i \dot{q}_j, \quad V = V(q),
\]

and the quadratic form \( T \) is positive definite. Consequently the Lagrange equations can be written in the form

\[
\sum_{j=1}^{s} \mu_{kj}(q) \ddot{q}_j + \sum_{i,j=1}^{s} \rho_{k,ij}(q) q_i \dot{q}_j + \frac{\partial V}{\partial q_k} = 0 \quad (1 \leq k \leq s)
\]

with certain coefficients \( \rho_{k,ij}(q) \). We can resolve this system with respect to \( \ddot{q}_j \) (\( 1 \leq j \leq s \)) since the matrix \( (\mu_{kj}) \) is non-singular. The equilibrium positions of the conservative mechanical system under consideration coincide with the stationary points of the potential energy:

\[
\frac{\partial V}{\partial q_k} = 0 \quad (1 \leq k \leq s).
\]

If \( q = 0 \) is an equilibrium position and if the motions which begin close to zero with small initial velocities remain as such for all time\(^{16} \), then it is natural to linearise the equations (14) for \( q, \dot{q} \):

\[
\sum_{j=1}^{s} \mu_{kj}(0) \ddot{q}_j + \sum_{j=1}^{s} \lambda_{kj} q_j = 0.
\]  

Here \( \lambda_{kj} = \left. \frac{\partial^2 V}{\partial q_k \partial q_j} \right|_{q=0} \). Equations (15) describe (approximately) small oscillations close to stable equilibrium. For \( s = 1 \) we have the harmonic oscillator, \( \mu \ddot{q}_1 + \lambda q_1 = 0 \), oscillating according to the rule \( q_1 = A \cos \omega t + B \sin \omega t \) with frequency \( \omega = \sqrt{\frac{\lambda}{\mu}} \). In the general case, because of the positive definiteness of the quadratic form \( T \), we can transform the system (15) by means of a non-singular linear substitution

\[
q_j = \sum_{l=1}^{s} c_{jl} Q_l \quad (1 \leq j \leq s)
\]

to the form

\(^{16}\)The assumption about the stability of the equilibrium position \( q = 0 \) which is contained in this sentence is valid if the potential energy \( V(q) \) has a strict local minimum at the point \( q = 0 \) (Lagrange-Dirichlet theorem).
\[
\ddot{Q}_l + \omega_l^2 Q_l = 0 \quad (1 \leq l \leq s),
\]
i.e. to system of \( s \) independent harmonic oscillators (normal modes) with frequency spectrum \( (\omega_l)_l^s \). The general solution of the system (15) is the superposition of these harmonic oscillations – this is a very simple example of harmonic analysis.

If the system is not conservative then in the Lagrange equations the right hand sides appear as non-potential generalised forces. These forces can act in such a way that the total energy of the system will be diminished (dissipation). In this case damped oscillations appear in the linear approximation.

We note that equations (15) become exact if we replace \( L \) by its second differential at the point \( q = 0 \).

Up to now we have considered ordinary differential equations describing the dynamics of a system with finitely many degrees of freedom. Systems with infinitely many degrees of freedom are usually described by partial differential equations. As an example let us consider small vibrations of a uniform string (with density \( \rho \)). In equilibrium let it occupy the interval \([0, l]\) of the \( x \)-axis and let it undergo a tension \( \lambda \) and then begin to oscillate (without friction) in a horizontal plane \( x, u \) so that at time \( t \) the point \((x, 0)\) is displaced to the point \((x, u(x, t))\) (transverse vibrations).

A satisfactory approximation to the action is provided by

\[
J = \frac{1}{2} \int_0^t \int_0^l (\rho u_t^2 - \lambda u_x^2) \, dt \, dx.
\]

This is a quadratic (in the obvious sense) functional on the space of smooth functions \( u(x, t) (0 \leq x \leq l, t_1 \leq t \leq t_2) \). With the assumption that \( u \in C^2 \) there follows from the variational principle \( \delta J = 0 \) the one-dimensional wave equation

\[
\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad \left( a = \frac{\lambda}{\sqrt{\rho}} \right).
\]

Equation (16) arises systematically in the description of oscillations of one-dimensional systems. Moreover in certain situations it can be exact (for example, in electrodynamics). For \(-\infty < x, t < \infty\) its general solution has the form

\[
u(x, t) = f(x + at) + g(x - at),
\]

where \( f, g \) are arbitrary functions (of class \( C^2 \)). Thus \( u(x, t) \) turns out to be the superposition of two travelling waves moving with speed \( a \) along the \( x \)-axis in opposite directions. Given the initial conditions

\[
u(x, 0) = \phi(x), \quad u_t(x, 0) = \psi(x)
\]

(i.e. the initial configuration of the string and the initial assignment of velocities)

---

The \( n \)-dimensional wave equation has form

\[
\frac{\partial^2 u}{\partial t^2} = a^2 \Delta u,
\]

where \( \Delta \) is the Laplace operator or Laplacian:

\[
\Delta u = \sum_{k=1}^n \frac{\partial^2 u}{\partial x_k^2}.
\]

For \( n = 2 \) this is the equation of small transverse vibrations of a uniform membrane.
D'Alembert's formula follows from (17):

\[ u(x, t) = \frac{\phi(x + at) + \phi(x - at)}{2} + \frac{1}{2a} \int_{x-at}^{x+at} \psi(\xi) \, d\xi. \]

For a finite string \((0 \leq x \leq l)\) boundary conditions must be given in addition, for example

\[ u(0, t) = 0, \quad u(l, t) = 0. \tag{19} \]

If moreover \(\phi \in C^2, \psi \in C^1\) and \(\phi, \phi'', \psi\) vanish at the ends of the interval \([0, l]\) (this is all necessary in order that \(u \in C^2\)), then the solution is found as before by D'Alembert's formula where \(\phi, \psi\), which are initially given for \(0 \leq x \leq l\), are then extended to the whole axis as odd periodic functions (with period \(2l\)).

We give one further classical partial differential equation – the heat equation for a bar\(^{18}\):

\[ \frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} \tag{20} \]

\((u = u(x, t)\) is the temperature at the point \(x\) at time \(t \geq 0)\). In order to extract a unique solution it is necessary to give the initial condition \(u(x, 0) = \phi(x)\) and, if the bar is finite, boundary conditions at each end such as (19).

\[ \bullet \frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} \]

§ 2. The Fourier Method and Related Questions

This method\(^1\) works effectively in a wide circle of problems of mathematical physics. Attempts to justify it in a sufficiently general form led to the creation of spectral theory, a most important branch of functional analysis. Eigenvalues (the spectrum in this or some more general sense) and eigenvectors of linear operators are the basic concepts of this theory. Finite-dimensional spectral theory is completely resolved by algebraic techniques but the infinite-dimensional situation requires analytical methods.

2.1. The Vibrations of a String. In this problem the physical significance of the Fourier method is the construction of the solution by means of the superposition of standing waves i.e. solutions of form \(u(x, t) = X(x)T(t)\). Separation of variables results from substitution in equation (16):

\[ \frac{X''}{X} = \frac{T''}{a^2T} = -\omega^2, \]

where \(\omega = \text{const}\). The boundary conditions (19) require that \(X(0) = 0, X(l) = 0\).

\(^{18}\)The \(n\)-dimensional heat equation has form \(\frac{\partial u}{\partial t} = \kappa \Delta u\).

\(^1\)Daniel Bernoulli and Euler were in possession of this method long before Fourier.
The frequency spectrum is determined from these to be \( \omega_n = \frac{n\pi}{l} \) \((n = 1, 2, 3, \ldots)\) and correspondingly
\[
X_n(x) = \sin \frac{n\pi}{l} x, \quad T_n(t) = A_n \cos \frac{n\pi a}{l} t + B_n \sin \frac{n\pi a}{l} t.
\]

The series
\[
\sum_{n=1}^{\infty} X_n(x) T_n(t)
\]
formally satisfies the equation, since it is linear, and clearly also the boundary conditions. The initial conditions (18) immediately reduce to the problem of expansion in trigonometric series:
\[
\phi(x) = \sum_{n=1}^{\infty} A_n \sin \frac{n\pi}{l} x, \quad \psi(x) = \sum_{n=1}^{\infty} B_n \frac{n\pi a}{l} \sin \frac{n\pi}{l} x.
\]

At one time a bitter controversy arose over this problem between Euler and D'Alembert, which was only resolved in the 19th century after the work of Fourier and after N.I. Lobachevskij (1835) and Dirichlet (1837) gave a more general definition of the idea of a function and established the first tests for the convergence of Fourier series. In the present context the Fourier series (21) of the function \( \phi \) is defined by the Euler-Fourier formulae for the Fourier coefficients
\[
A_n = \frac{2}{l} \int_0^l \phi(x) \sin \frac{n\pi}{l} x \, dx,
\]
which arise formally from the orthogonal system of functions \( \sin \frac{n\pi}{l} x \) on the interval \([0, l]\). Likewise
\[
B_n = \frac{2}{n\pi a} \int_0^l \psi(x) \sin \frac{n\pi}{l} x \, dx.
\]

With this the procedure for solving the problem on the vibrations of a finite string is formally completed.

**Theorem.** If \( \phi \in C^2[0, l], \psi \in C^1[0, l] \) and \( \phi, \phi'', \psi \) vanish at the points \( x = 0, l \), then the series
\[
\sum_{n=1}^{\infty} \left( A_n \cos \frac{n\pi a}{l} t + B_n \sin \frac{n\pi a}{l} t \right) \sin \frac{n\pi}{l} x
\]
with coefficients defined by formulae (22), (23), converges uniformly on the strip \( 0 \leq x \leq l, -\infty < t < \infty \) to a function \( u(x, t) \) of class \( C^2 \) which satisfies the equation for vibrations of a string, the boundary conditions \( u(0, t) = 0, u(l, t) = 0 \) and the initial conditions \( u(x, 0) = \phi(x), u_t(x, 0) = \psi(x) \).

---

2 The formulae are first encountered not in the work of these authors but in that of Clairaut (1754).
3 If the expansion (21) for \( \phi \) converges uniformly on \([0, l]\) then formulae (22) necessarily hold.
2.2. Heat Conduction. For a solution of the form \( u(x, t) = X(x)T(t) \) we obtain from equation (20):

\[
\frac{X''}{X} = \frac{T'}{\kappa T} = -\omega^2.
\]

If \( 0 \leq x \leq l \) then under the boundary conditions (19) we obtain the same spectrum as in the previous problem and the same \( X_n(x) \), but now

\[
T_n(t) = e^{-\kappa(n\pi/l)^2 t},
\]

which gives

\[
u(x, t) = \sum_{n=1}^{\infty} A_n e^{-\kappa(n\pi/l)^2 t} \sin \frac{n\pi}{l} x
\]

as the formal solution of the problem with initial condition \( u(x, 0) = \phi(x) \), the \( A_n \) being defined by the formulae (22).

Theorem. If \( \phi \in C^2[0, l] \) and \( \phi, \phi'' \) vanish at the points \( x = 0, l \), then the series (24) converges uniformly on the half-strip \( 0 \leq x \leq l, t \geq 0 \) to the solution of the heat equation with boundary conditions \( u(0, t) = 0, u(l, t) = 0 \) and initial condition \( u(x, 0) = \phi(x) \).

In the problems considered above, the solution \( u(x, t) \) is obtained in the form of a series in the eigenfunctions \( \sin \frac{n\pi}{l} x \) of the operator \( \frac{d^2}{dx^2} \) on \( [0, l] \) with the conditions \( u|_{x=0} = 0, u|_{x=l} = 0 \).

The essence of the Fourier method is the expansion in terms of the eigenfunctions of a linear differential operator (a sufficiently complete description of this class of operators will be given in the sequel).

2.3. The Classical Theory of Fourier Series. For simplicity let us assume in (21) that \( l = \pi \). If the series (21) for \( \phi \) (with \( l = \pi \)) converges to it everywhere on \( [0, \pi] \) then it converges on the whole axis to the function \( \tilde{\phi} \) obtained from \( \phi \) by extending it as an odd function to \( [-\pi, 0] \) and then as a periodic function (with period \( 2\pi \)) to the whole axis. If \( f \) is any (not necessarily odd) periodic function with period \( 2\pi \) which belongs to \( L^1(-\pi, \pi) \), then its Fourier series is defined as

\[
\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),
\]

where

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx;
\]

these are its Fourier coefficients.

*We can say that \( f \) is an \( L^1 \)-function on the unit circle \( T \).
For complex $f$ the complex form of the Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{in\xi}$$

(25')
is used more frequently; to transfer to this we have to take

$$c_n = \frac{1}{2} (a_n - ib_n \text{sgn } n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \, dx$$

($c_{-n} = \bar{c}_n$ for real $f$). Moreover the partial sums of the series (25') may be written in the form

$$S_n(x) = \sum_{|n| \leq n} c_n e^{inx}.$$  

The system $(e^{inx})_{n \in \mathbb{Z}}$ is orthonormal on $T$ if the measure of $T$ is normalised to 1.

A fundamental problem is to investigate the convergence of the Fourier series of a function $f$ to itself or, in a more general form, to reconstruct the function $f$ from its Fourier series. For example, in the proofs of the theorems in Sections 2.1, 2.2 use is made of the result that the Fourier series of a smooth function converges uniformly to it. The following is a more subtle test.

**Dirichlet-Jordan Theorem.** If $f$ is a continuous function of bounded variation then its Fourier series converges to it uniformly.

The proof of this theorem relies on the representation of the partial sums in the form

$$S_m(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_m(x - \xi) f(\xi) \, d\xi,$$  

(26)

where $D_m(x)$ is the Dirichlet kernel:

$$D_m(x) = \sum_{|n| \leq m} e^{inx} = \frac{\sin(m + \frac{1}{2})x}{\sin \frac{x}{2}}.$$  

(27)

It possesses the following basic properties:

1) $\frac{1}{2\pi} \int_{-\pi}^{\pi} D_m(x) \, dx = 1$;
2) if a function $f \in L^1$ vanishes on some neighbourhood of zero then

$$\lim_{m \to \infty} \int_{-\pi}^{\pi} D_m(x) f(x) \, dx = 0.$$  

We note that in formula (26) there appears the important convolution construction for two summable functions on the circle:

$$(g * f)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(x - \xi) f(\xi) \, d\xi.$$  

For the convolution of functions the corresponding Fourier coefficients are multiplied together (Convolution Theorem).
Property 2) follows from the Riemann-Lebesgue theorem: if \( h \in L^1 (-\infty, \infty) \) then

\[
\lim_{|\lambda| \to \infty} \int_{-\infty}^{\infty} h(x) e^{i\lambda x} \, dx = 0.
\]

In turn, the principle of localisation follows from property 2): the convergence and the sum of the Fourier series of a function \( f \in L^1 \) at a point \( x \) depends only on the germ of the function at this point.

The insufficiency of continuity alone even for the convergence of the Fourier series at each point was first shown in 1876 by Du Bois-Reymond in an example, after which Lebesgue (1905) constructed a continuous function whose Fourier series converges everywhere but not uniformly. This line of development, which compromised pointwise convergence as a tool for reconstructing a function from its Fourier series, was brought to a definite conclusion when A.N. Kolmogorov (1926) produced an example of a function in \( L^1 \) whose Fourier series diverges everywhere. In another direction, N.N. Luzin raised the question of the convergence almost everywhere of the Fourier series of any function \( f \in L^2 \). This question was resolved positively by Carleson (1966) and shortly afterwards Hunt extended Carleson's result to all \( L^p \) \( (p > 1) \).

A natural method of reconstructing a function from its Fourier series is provided by Cesàro summation

\[
\hat{S}_k(x) = \frac{1}{k+1} \sum_{m=0}^{k} S_m(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_k(x - \xi) f(\xi) \, d\xi, \tag{28}
\]

where \( F_k(x) \) is the Fejér kernel:

\[
F_k(x) = \frac{1}{k+1} \sum_{m=0}^{k} D_m(x) = \frac{\sin^2 \frac{k+1}{2} \pi x}{(k+1) \sin^2 \frac{x}{2}}
\]

\[
= \sum_{|n| \leq k} \left( 1 - \frac{|n|}{k+1} \right) e^{i nx}, \tag{29}
\]

the \( \hat{S}_k(x) \) are called the Fejér averages of the function \( f \).

**Lebesgue-Fejér Theorem.** For any function \( f \in L^1 \) the Fejér averages converge to \( f(x) \) almost everywhere\(^6\). If \( f \) is continuous then its Fejér averages converge to \( f \) uniformly.

The proof of this theorem is based on the representation (28) and the following basic properties of the Fejér kernel:

\(^6\) At all Lebesgue points, i.e. points \( x \) such that

\[
\lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{\epsilon}^{\epsilon} |f(x + \xi) - f(x)| \, d\xi = 0.
\]
1) \( \frac{1}{2\pi} \int_{-\pi}^{\pi} F_n(x) \, dx = 1; \)

2) \( \lim_{k \to \infty} F_k(x) = 0 \) uniformly on \( [-\frac{\pi}{2}, \pi] \) outside of any neighbourhood of zero;

3) \( F_k(x) \geq 0 \) \( (-\infty < x < \infty) \).

As compared with the Dirichlet kernel, 3) is the most significant property and it leads to quite different behaviour of the Fejér averages from that of the partial sums of the Fourier series.

**Corollary 1.** A function \( f \in L^1 \) is uniquely determined by its Fourier series.

This justifies the usual notation

\[
f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}
\]

for the correspondence between functions and their Fourier series (and the analogous notation in the real form).

**Corollary 2** (Theorem of Weierstrass). Any continuous periodic function (with period \( 2\pi \)) is the uniform limit of a certain sequence of trigonometric polynomials

\[
T_m(x) = \sum_{|n| \leq m} c_{m,n} e^{inx}.
\]

The theory of Fourier series for functions of class \( L^2 \) turns out to be particularly satisfactory. The reason for this is the Euclidean geometry and completeness of the space \( L^2 \). Moreover the \( L^2 \)-theory is most important for many applications, which is explained in particular by the physical significance of the integral of the square of the modulus. For example, the total energy of a string (in the approximation for small vibrations) is equal to

\[
E = \frac{1}{2} \int_0^l (\rho u_t^2 + \lambda u_x^2) \, dx;
\]

in quantum mechanics the integral of the square of the modulus of the wave function \( \psi(x, t) \) over any volume \( \Omega \) is the probability of finding particles in this volume and so on.

In the \( L^2 \)-theory it is necessary to consider for \( f \in L^2 \) the quantity

\[
E_m(f) = \| f - S_m \|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - S_m(x)|^2 \, dx.
\]

The following expression for \( E_m(f) \) is easily obtained:

\[
E_m(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx - \sum_{|n| \leq m} |c_n|^2,
\]

from which follows immediately Bessel’s inequality

\[
\sum_{n=-\infty}^{\infty} |c_n|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx
\]

and, consequently,
Further for any trigonometric polynomial (30)

\[ \|f - T_m\|^2 = \|f - S_m\|^2 + \|S_m - T_m\|^2 \geq E_m(f) \]  

(32)

(here Pythagoras's theorem can be applied because of the orthogonality of the function \( f - S_m \) to all \( e^{inx} (|n| \leq m) \) and therefore to all trigonometric polynomials of degree \( \leq m \), in particular, \( S_m - T_m \)). The inequality (32) shows that the \( m \)th partial sum of the Fourier series of the function \( f \) (and only it) gives the best least squares approximation to \( f \) by means of trigonometric polynomials of degree \( \leq m \) (the error of this approximation is equal to \( \sqrt{E_m(f)} \)). If \( f \) is continuous we can assume that \( T_m(x) \) converges uniformly to \( f \) as \( m \to \infty \) and moreover in \( L^2 \). The last assertion is correct for any \( f \in L^2 \), so that \( f \) can be approximated in \( L^2 \) to any degree of accuracy by continuous functions. But then from (32) \( E_m(f) \to 0 \) as \( m \to \infty \); thus we have

**Theorem.** For any \( f \in L^2 \) its Fourier series converges to \( f \) in \( L^2 \), i.e. in mean square.

In this case Parseval's equality\(^7\) holds (see (31)):

\[ \sum_{n=-\infty}^{\infty} |c_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx. \]

On the other hand the Riesz-Fischer theorem on the completeness of the space \( L^2 \) implies the convergence in \( L^2 \) of any trigonometric series

\[ \sum_{n=-\infty}^{\infty} c_n e^{inx} \]  

(33)

which satisfies the condition

\[ \sum_{n=-\infty}^{\infty} |c_n|^2 < \infty. \]  

(34)

But then (33) will be the Fourier series of its sum. Thus we have a one-to-one correspondence \( \mathcal{F} \) between the function space \( L^2(-\pi, \pi) \) and the space of scalar (complex) sequences \( (c_n)_{-\infty}^{\infty} \) which satisfy condition (34). The latter space, which is usually denoted by\(^8\) \( l^2 \), is \( L^2(\mathbb{Z}, \mu) \) where \( \mu \) is the measure on \( \mathbb{Z} \) such that\(^9\) \( \mu(\{n\}) = 1 \) for all integers \( n \). Parseval's equality says that \( \mathcal{F} \) is an isometry. Moreover \( \mathcal{F} \) is an isomorphism of the linear spaces \( L^2(-\pi, \pi) \) and \( \ell^2 \). It is

---

\(^7\) Although Parseval wrote down this equality away back in 1799, it was only proved considerably later.

\(^8\) The notation \( l^2 \) is also used for the space of one-sided numerical (complex) sequences, the squares of the moduli of whose terms form convergent series (we also retain it when we restrict to real numbers). The spaces \( l^p \) \((p \geq 1)\) are defined similarly.

\(^9\) If \( \mu(\{n\}) = \mu_n \), where \( (\mu_n) \) is a given (weight) sequence of positive numbers, then we use the notation \( l^p_\mu \) \((p \geq 1)\).
appropriate to note here that from Parseval's equality follows the more general identity

\[ \sum_{n=-\infty}^{\infty} c_n \bar{d}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) g(x) \, dx, \]

where not only \( f \) but also \( g \) belong to \( L^2 \), the \( d_n \) being the Fourier coefficients of \( g \). This means geometrically that \( \langle \mathcal{F} f, \mathcal{F} g \rangle = \langle f, g \rangle \).

The problem of absolute convergence occupies a significant place in the theory of trigonometric series. In the case of a series of the form (25') absolute convergence everywhere is clearly equivalent to \( \sum_n |c_n| < \infty \). In the case of series (25) it is equivalent to

\[ \sum_n (|a_n| + |b_n|) < \infty, \]  

but this is no longer obvious (in one direction) and constitutes the Denjoy-Luzin theorem: if the series

\[ \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \]

is absolutely convergent on a set of positive measure then (35) holds. In contrast to this we have the example of the series

\[ \sum_{n=1}^{\infty} \sin(n!x), \]

which terminates for each \( x \) that is commensurable with \( \pi \).

If a trigonometric series converges everywhere then by (35) it converges uniformly, its sum \( f \) is continuous and it is the Fourier series of \( f \). Conversely if a continuous function \( f \) is given then the absolute convergence of its Fourier series depends essentially on the behaviour of its modulus of continuity

\[ \omega(\delta) = \max_{|x-y| \leq \delta} |f(x) - f(y)| \]

as \( \delta \to 0 \). We say that \( f \) belongs to the Lipschitz class \( \text{Lip} \alpha \) with exponent \( \alpha (0 < \alpha \leq 1) \) and write \( f \in \text{Lip} \alpha \) if \( \omega(\delta) = O(\delta^\alpha) \). S.N. Bernshtein proved that if \( f \in \text{Lip} \alpha \) for some \( \alpha > \frac{1}{2} \) then its Fourier series is absolutely convergent\(^{11}\), while counterexamples are to be found in Lip \( \frac{1}{2} \).

The functions with absolutely convergent Fourier series form not only a linear space but also an algebra, for if

\[ f_1(x) = \sum_{n=-\infty}^{\infty} c_n^{(1)} e^{inx}, f_2(x) = \sum_{n=-\infty}^{\infty} c_n^{(2)} e^{inx} \]

and \( f = f_1 f_2 \), then

\(^{10}\) Or Hölder class.

\(^{11}\) A more general condition is \( \int_0^\infty \omega(\delta) \delta^{-2\alpha} \, d\delta < \infty \) (\( \alpha > 0 \)).
\[ f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \]

where
\[ c_n = \sum_{k=-\infty}^{\infty} c_{n-k}^{(1)} c_{k}^{(2)} \quad (n = 0, \pm 1, \pm 2, \ldots). \]

the convolution of the sequences \((c_n^{(1)}), (c_n^{(2)})\), and
\[ \sum_{n=-\infty}^{\infty} |c_n| \leq \sum_{k=-\infty}^{\infty} |c_k^{(1)}| \sum_{l=-\infty}^{\infty} |c_l^{(2)}|. \]

This is the Wiener algebra \( W \), so-called because of its connection with the following theorem.

**Theorem of Wiener.** If \( f \in W \) and \( f(x) \) is nowhere equal to zero, then \( \frac{1}{f} \in W \).

The algebra \( W \) is a very interesting and significant example of a commutative Banach algebra. It is now customary to apply this term to a Banach space which has the structure of an associative, commutative algebra with identity \( e \) and whose norm has the following properties relating to this structure:
\[ \| f_1 f_2 \| \leq \| f_1 \| \cdot \| f_2 \|; \quad \| e \| = 1. \]

The theory of commutative Banach algebras (under the name commutative normed rings) was originated by I.M. Geland but the theorem of Wiener served as its first touchstone. We note that the natural norm on the algebra \( W \) is \( \| f \| = \sum \| c_n \| \). In other words, the Fourier series expansion transforms \( W \) isometrically and isomorphically into the Banach algebra \( L^1 \) with convolution as multiplication.

**2.4. General Orthogonal Series.** The purely geometric character of the \( L^2 \)-theory of Fourier series was first realised by Hilbert who considered in place of the system \((e^{inx})\) in \( L^2(-\pi, \pi) \) an arbitrary orthogonal system \((\phi_n(x))\) in \( L^2(a, b) \); it is even possible to consider such a system in \( L^2(X, \mu) \) for any space \( X \) with some measure \( \mu \), since this in no way affects the argument. Defining the generalised Fourier coefficients as
\[ c_n = \frac{1}{\| \phi_n \|^2} \int_X f \overline{\phi_n} \, d\mu = \frac{(f, \phi_n)}{\| \phi_n \|^2}, \]
we can introduce for any function \( f \in L^2(X, \mu) \) its generalised\(^{12} \) Fourier series \( \sum_n c_n \phi_n(x) \). The system \((\phi_n)\) is said to be complete if it is possible to approximate any \( f \in L^2(X, \mu) \) in mean square to any degree of precision by means of linear combinations of functions of the system\(^{13} \). However, as before, the partial sums

\(^{12}\)The word 'generalised' will be dropped in the sequel.

\(^{13}\)In this definition orthogonality is not necessary. We can also discuss the completeness of systems in other topologies, in particular in \( L^p(X, \mu) \) \((p \geq 1)\).
of the Fourier series are the best approximations of this type. The property of being complete for the system \( (\phi_n) \) is therefore equivalent to the fact that for any \( f \in L^2(X, \mu) \) its Fourier series converges to \( f \) in mean square. Even if the system is not complete, the Fourier series of all functions in \( L^2(X, \mu) \) converge in mean square because of Bessel's inequality

\[
\sum_n |c_n|^2 \|\phi_n\|^2 \leq \int_X |f|^2 \, d\mu
\]

and the Reisz-Fischer theorem. In the case of a complete system (and only in this case) the functions \( f \in L^2(X, \mu) \) are uniquely determined by their Fourier coefficients. In order that the series converge in \( L^2(X, \mu) \) to \( f \), it is necessary and sufficient that Parseval's equality should hold:

\[
\sum_n |c_n|^2 \|\phi_n\|^2 = \int_X |f|^2 \, d\mu.
\]

The system \( (\phi_n) \) is therefore complete if and only if Parseval's equality holds for all \( f \in L^2(X, \mu) \). In this case the mapping \( f \mapsto (c_n) \) turns out to be an isometric isomorphism of the spaces \( L^2(X, \mu) \) and \( l^2 \) where \( v_n = \|\phi_n\|^2 \).

### 2.5. Orthogonal Polynomials.

Suppose there is given on some interval \( I \) of the real axis a non-decreasing function \( \sigma(x) \) such that all the powers \( x^k \) \( (k = 0, 1, 2, \ldots) \) belong to \( L^2_\sigma \). Suppose that the set of points of increase of the function \( \sigma \) is infinite. Applying the process of orthogonalisation (with normalisation\(^{14}\)) to the system of powers we obtain the system of orthogonal polynomials \( (P_k(x))_{k=0}^{\infty} \) (deg \( P_k = k \)), which is uniquely defined for the given function \( \sigma \) (or measure \( \mu_\sigma \)). If \( \sigma \) is absolutely continuous its derivative \( p = \sigma' \) is called a weight and then we speak of polynomials which are orthogonal with respect to the given weight \( p \). The classical examples, which arise in problems of mathematical physics, are:

1. \( p(x) = e^{-x^2} \) \((- \infty < x < \infty)\) – Hermite polynomials;
2. \( p(x) = e^{-x} \) \((0 < x < \infty)\) – Laguerre polynomials;
3. \( p(x) = 1 \) \((- 1 \leq x \leq 1)\) – Legendre polynomials.

All three systems are complete in the corresponding \( L^2_\sigma \). In other words, in each of these spaces the system of powers is complete. As a result of Weierstrass's theorem on the uniform approximation of continuous functions by means of polynomials, the system of powers is complete on a finite interval for any \( \sigma \). On an infinite interval whether it is complete depends on the weight. For example, if \( p(x) = e^{-x^2|x|} \) then the system is not complete since

\[
\int_{-\infty}^{\infty} x^k e^{-x^2|x|} \sin(2\pi \ln |x|) \, dx = 0 \quad (k = 0, 1, 2, \ldots)
\]

(any function which has arbitrary close mean-square approximations by means of polynomials for the given weight is orthogonal to \( \sin(2\pi \ln |x|) \)). This example of Stieltjes is covered by the following test (M.G. Krejn, 1945).

\(^{14}\) It is possible to carry it out so that the highest coefficients are positive; this is assumed in the sequel.
Theorem. If \( \sigma'(x) \) is the derivative of the absolutely continuous part of a function \( \sigma(x) (-\infty < x < \infty) \) and \( \int_{-\infty}^{\infty} \frac{\ln \sigma'(x)}{1 + x^2} \, dx > -\infty \), then the system of powers is not complete in \( L^2_\sigma \).

The converse assertion is true if \( \sigma(x) \) is absolutely continuous and the weight \( p(x) \) satisfies the conditions:
1) sup \( p(x) < \infty \); 2) \( p(-x) = p(x) \);
3) for \( x > 0 \) the function \( -\ln(1 + x^2)p(x) \) is non-decreasing and convex relative to \( \ln x \).

Thus if the conditions listed still hold and
\[
\int_{-\infty}^{\infty} \frac{\ln p(x)}{1 + x^2} \, dx = -\infty,
\]
then the system of powers is complete in \( L^2_\sigma \).

The following general criterion for completeness was given by B.Ya. Levin.

Theorem. Let \( \mathcal{M}_\sigma \) be the set of all polynomials satisfying the condition
\[
\int_{-\infty}^{\infty} \frac{|Q(x)|^2}{1 + x^2} \, dx \leq 1
\]
and put
\[
M(x) = \sup_{Q \in \mathcal{M}_\sigma} |Q(x)|.
\]
In order that the system of powers be complete in \( L^2_\sigma \) it is necessary and sufficient that
\[
\int_{-\infty}^{\infty} \frac{\ln M(x)}{1 + x^2} \, dx = \infty.
\]

This theorem is related to the criteria of N.I. Akhiezer and S.N. Bernshtein (1953) and S.N. Mergelyan (1954), relating to weighted uniform approximation by polynomials on the whole axis. In his lectures (1957) B.Ya. Levin developed a general theory which covers weighted approximation both in the uniform case and in \( L^2_\sigma \).

2.6. The Power Moment Problem. Let us denote by \( I \) any one of the three intervals \( (-\infty, \infty), [0, \infty), [0, 1] \). Let \( \sigma(x) (x \in I) \) be a non-decreasing function and as in Section 2.5 suppose
\[
\int_I |x|^k \, d\sigma(x) < \infty
\]
(in the case of \( I = [0, 1] \) this condition follows from the boundedness of \( \sigma \)). The numbers
\[
m_k = \int_I x^k \, d\sigma(x) \quad (k = 0, 1, 2, \ldots)
\]
are called the power moments of the function \( \sigma \) (measure \( \mu_\sigma \)). If we interpret the measure \( \mu_\sigma \) as a mass distributed on \( I \) then \( m_0 \) is the mass of the whole \( \mathbb{R}^d \), \( m_1 \) is the static moment, \( m_2 \) is the moment of inertia relative to the point \( x = 0 \). The problem is to give an intrinsic characterisation of the class of moment sequences \( (m_k)_k \). For any \( I \) an obvious necessary condition is the positivity of the Hankel matrix \( M = (m_{k+j})_{k,j=0}^\infty \):

\[
\sum_{k=0}^n m_{k+j} \xi_k \xi_j \geq 0 \quad \text{for all } n = 0, 1, 2, \ldots \quad \text{and for all } (\xi_1, \xi_2, \ldots, \xi_n) \in \mathbb{R}^n.
\]

This condition is also sufficient in the case \( I = (-\infty, \infty) \) (theorem of Hamburger). In the case \( I = [0, \infty) \) necessary and sufficient conditions are provided by positivity of the matrices \( M \) and \( M_1 = (m_{k+j+1})_{k,j=0}^\infty \); in the case \( I = [0, 1] \) we have an analogous (but more cumbersome) criterion, although the following theorem is of greater interest.

**Theorem of Hausdorff.** In order that \( (m_k) \) be a moment sequence on \([0, 1]\) it is necessary and sufficient that the inequalities

\[
\Delta^n_{\xi} m_k \equiv \sum_{j=0}^n (-1)^j \binom{n}{j} m_{k+j} \geq 0 \quad (n, k = 0, 1, 2, \ldots)
\]

hold.

As before the necessity is trivial. The proof of the sufficiency can be carried out on the basis of the uniform approximation of a function \( f \in C[0, 1] \) by means of the Bernshtein polynomials

\[
B_n(x; f) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1 - x)^{n-k}.
\]

This tool was proposed at one time by S.N. Bernshtein for the proof of Weierstrass theorem. At the same time we obtain from Weierstrass theorem the definiteness of the moment problem on an interval, i.e. the uniqueness of the corresponding measure. As the example of Stieltjes shows, definiteness fails in general on the semiaxis or the whole axis. Carleman (1922) found the following sufficient condition for definiteness of the moment problem on \([0, \infty)\):

\[
\sum_{k=0}^\infty \sqrt{\frac{m_k}{m_{k+1}}} = \infty.
\]

It is satisfied for example by \( m_k = k! \) In this case the unique (up to an additive constant) solution of the moment problem is \( \sigma(x) = 1 - e^{-x} \).

On the whole axis an analogous sufficient condition is as follows:

\[
\sum_{k=0}^\infty \sqrt{\frac{m_{2k}}{m_{2k+2}}} = \infty.
\]

The sequence \( m_{2k} = \Gamma(k + \frac{1}{2}) \) along with \( m_{2k+1} = 0 \ (k = 0, 1, 2, \ldots) \), which cor-
responds to the weight \( p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \), provides an example where this condition is satisfied.

The investigation of the moment problem is closely connected with the properties of orthogonal polynomials. Let us consider the moment problem on the whole axis for the sequence \( (m_k)_0^\infty \). We introduce on the linear space of real polynomials the linear functional

\[
\Theta(R) = \sum_{k=0}^{n} m_k r_k \quad \left( R(x) = \sum_{k=0}^{n} r_k x^k \right).
\]

If the given moment problem is solvable and \( \sigma(x) \) is any solution of it, then

\[
\Theta(R) = \int_{-\infty}^{\infty} R(x) \, d\sigma(x).
\] (36)

Thus, all solutions generate one and only one linear functional and one and only one scalar product \( (P, Q) = \Theta(PQ) \) (it is important for this construction that the polynomials form not only a linear space but also an algebra). Positivity of this scalar product follows from positivity of the functional \( \Theta \), while for its non-degeneracy it is necessary and sufficient that the set of points of increase of the function \( \sigma \) be infinite, which in turn is equivalent to the rank of the matrix \( M \) being infinite – this will be assumed from now on. It follows from what has been said that the system of orthogonal polynomials \( (P_k(x))_0^\infty \) is uniquely defined by the moments \( (m_k)_0^\infty \) even in the indefinite case. We shall assume further that \( m_0 = 1 \) and then \( P_0(x) \equiv 1 \). The theory set out below was established in 1922 in the works of Hellinger, Nevanlinna and M. Riesz and was subsequently developed further by H. Weyl (1935).

**Lemma.** If the series \( \sum_k |P_k(z)|^2 \) converges even at one point \( z \) (im \( z \neq 0 \)) then it converges uniformly on each bounded region.

In all cases the quantity

\[
r(z) = \frac{1}{2|\text{im} \, z|} \left( \sum_k |P_k(z)|^2 \right)^{-1}
\]

is called the Weyl radius at the point \( z \). The lemma means that on the set \( \text{im} \, z \neq 0 \) the Weyl alternative holds: either \( r(z) = 0 \) for all \( z \) or \( r(z) > 0 \) for all \( z \).

Let us introduce the system of polynomials

\[
Q_k(z) = \Theta_x \left( \frac{P_k(z) - P_k(x)}{z - x} \right) \quad (k = 0, 1, 2, \ldots),
\] (37)

associated with \( (P_k)_0^\infty \). If we put

\[
w_\sigma(z) = \int_{-\infty}^{\infty} \frac{d\sigma(x)}{x - z} \quad (\text{im} \, z \neq 0),
\]

then as a result of (36) we have
\[ Q_k(z) = -w_\sigma(z)P_k(z) + c_k(z), \]

where the \( c_k(z) \) are the Fourier coefficients of the function \( (x - z)^{-1} \). Bessel inequality gives

\[
\sum_{k=0}^{\infty} |w_\sigma(z)P_k(z) + Q_k(z)|^2 \leq \int_{-\infty}^{\infty} \frac{d\sigma(x)}{|x - z|^2} = \frac{\text{im} w_\sigma(z)}{\text{im} z}. \tag{38}
\]

Consequently the point \( w_\sigma(z) \) belongs to the set \( W(z) \) defined by the inequality

\[
\sum_{k=0}^{\infty} |wP_k(z) + Q_k(z)|^2 \leq \frac{\text{im} w}{\text{im} z}.
\]

This is a disk of radius \( r(z) \) for \( r(z) > 0 \) and a point for \( r(z) = 0 \). Correspondingly \( W(z) \) is called the Weyl disk or point.

**Theorem 1.** *The moment problem is definite in the point case and indefinite in the disk case.*

In fact, if in the point case there are two solutions \( \sigma_1(x), \sigma_2(x) \) then \( w_{\sigma_1}(z) = w_{\sigma_2}(z) \) for \( \text{im} z \neq 0 \), i.e.

\[
\int_{-\infty}^{\infty} \frac{d\sigma_1(x)}{x - z} = \int_{-\infty}^{\infty} \frac{d\sigma_2(x)}{x - z},
\]

from which it follows that \( \sigma_1(x) \) essentially coincides with \( \sigma_2(x) \) (see Section 1.5).

In the disk case for each point \( \zeta \in W(z) \) there exists a solution \( \sigma(x) \) of the moment problem such that \( w_\sigma(z) = \zeta \). A solution \( \sigma(x) \) is called \( z \)-extremal if the point \( w_\sigma(z) \) lies on the boundary of the disk, i.e. equality is attained in (38) or in other words, Parseval equality holds for \( (x - z)^{-1} \). Therefore, if the system of powers is complete in \( L^2_\sigma \), then \( \sigma(x) \) is \( z \)-extremal for all \( z \) (\( \text{im} z \neq 0 \)). The converse is true even in a strengthened form.

**Theorem 2.** *If the solution \( \sigma(x) \) is \( z \)-extremal for some \( z \) (\( \text{im} z \neq 0 \)) then the system of powers is complete in \( L^2_\sigma \).*

Since this solution is \( z \)-extremal for all \( z \), we can call it simply extremal. The unique solution in the point case is referred to in the same way. Analogous to Theorem 2 we have the following theorem.

**Theorem 3.** *If the moment problem has a unique solution \( \sigma(x) \) then the system of powers is complete in \( L^2_\sigma \).*

The description of all solutions of the moment problem in the indefinite case is of considerable interest. Some further concepts are required for its presentation.

We say that a function \( u(z) \) which is analytic on the half-plane \( \text{im} z > 0 \) belongs to the class of Nevalinna (in short – the class \( N \)) if \( \text{im} u(z) \geq 0 \). For example, any function \( w_\sigma(z) \) is of this type.

**Theorem 4.** *In order that the function \( u(z) \) (\( \text{im} z > 0 \)) be in the class \( N \) it is necessary and sufficient that it permits an integral representation.*
where \( u(z) \) is a non-decreasing bounded function, \( \alpha \geq 0 \) and \( \beta \) is real.

**Corollary.** In order that the function \( w(z) (im z > 0) \) be of the form

\[
w(z) = \int_{-\infty}^{\infty} \frac{d\tau(x)}{x - z},
\]

where \( \tau(x) \) is a non-decreasing bounded function, it is necessary and sufficient that \( w(z) \) belongs to the class \( N \) and satisfies the condition

\[
\lim_{y \to +\infty} |yw(iy)| < \infty.
\]

In the indefinite case the series \( \sum_k Q(z)^2 \) converges uniformly in each bounded region. Let us put

\[
a_{11}(z) = z \sum_k Q_k(0)Q_k(z), \quad a_{12}(z) = 1 + z \sum_k P_k(0)Q_k(z),
\]

\[
a_{21}(z) = -1 + z \sum_k Q_k(0)P_k(z), \quad a_{22}(z) = z \sum_k P_k(0)P_k(z).
\]

This defines the Nevalinna matrix

\[
N(z) = \begin{pmatrix} a_{11}(z) & a_{12}(z) \\ a_{21}(z) & a_{22}(z) \end{pmatrix},
\]

an entire matrix-function of the complex variable \( z \), which is real on the real axis and identically unimodular, i.e. \( \det N(z) = 1 \).

**Theorem 5.** The formula

\[
\int_{-\infty}^{\infty} \frac{d\sigma(x)}{x - z} = -\frac{a_{11}(z)u(z) - a_{12}(z)}{a_{21}(z)u(z) - a_{22}(z)}
\]

establishes a \( 1 - 1 \) correspondence between the set of solutions of the moment problem and the set of functions of class \( N \) augmented by the constant \( \infty \). Under this correspondence the set of extremal solutions is mapped onto the set of real constants (\( \infty \) included).

**Corollary.** If \( \sigma(x) \) is an extremal solution of the moment problem then \( w_\sigma(z) \) is a meromorphic function.

Its poles are clearly real and coincide with the points of increase of the function \( \sigma(x) \).

\[\text{It is of zero exponential type: } \lim_{|z| \to \infty} \frac{\ln \max_{i,x,k}|a_{ik}(z)|}{|z|} = 0.\]
2.7. Jacobian Matrices. If \( j \leq k - 2 \) the polynomial \( xP_k(x) \), being a linear combination of the polynomials \( P_0(x), P_1(x), \ldots, P_{k-1}(x) \), is orthogonal to \( P_j(x) \). But then \( xP_k(x) \) is orthogonal to \( P_j(x) \). Consequently

\[
xP_k(x) = a_k P_k(x) + b_k P_{k+1}(x) + \beta_k P_{k-1}(x) \quad (k = 0, 1, 2, \ldots)
\]

(39)

Here

\[
P_{-1}(x) \equiv 0, \quad a_k = \langle xP_k(x), P_k(x) \rangle, \quad b_k = \langle xP_k(x), P_{k+1}(x) \rangle > 0,16
\]

\[
\beta_k = \langle xP_k(x), P_{k-1}(x) \rangle = \langle xP_{k-1}(x), P_k(x) \rangle = b_{k-1} \quad (b_{-1} = 0).
\]

The recurrence formula (39) with \( k \geq 1 \) defines the sequence \( (P_k(x))_0^\infty \) under the initial conditions \( P_0(x) = 1, P_1(x) = \frac{x - a_0}{b_0} \). By means of (37) this formula defines the associated sequence \( (Q_k(x))_0^\infty \) under the initial conditions \( Q_0(x) = 0, Q_1(x) = \frac{1}{b_0} \).

The infinite matrix

\[
J = \begin{bmatrix}
    a_0 & b_0 & 0 & 0 & \ldots \\
    b_0 & a_1 & b_1 & 0 & \ldots \\
    0 & b_1 & a_2 & b_2 & \ldots \\
    \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}
\]

(40)

is called the Jacobian matrix of the given moment sequence \( (m_k)_0^\infty \) (or of the given \( \sigma(x) \)). We remark that \( J \) uniquely determines the moments since from the expansion

\[
x^k = \sum_{j=0}^{k} c_{kj} P_j(x)
\]

there follows \( m_k = c_{k0} \) \( (k = 0, 1, 2, \ldots) \). It can be shown that for any real matrix of the form (40) with positive off-diagonal elements the sequence of polynomials \( (P_k(x))_0^\infty \) defined by it is orthogonal with respect to some \( \sigma(x) (-\infty < x < \infty) \).

We shall consider the matrix \( J \) as a linear operator on the (infinite-dimensional) linear space of sequences \( (v_k)_0^\infty \) of complex numbers. The recurrence formula (39) (with the replacement of \( x \) by an arbitrary \( z \in \mathbb{C} \)) shows that the sequence \( \Pi(z) = (P_k(z))_0^\infty \) is an eigenvector for \( J \) with corresponding eigenvalue \( z \). Clearly, apart from multiples, it is unique in this property, i.e. all eigenspaces of the operator \( J \) are one-dimensional. The Weyl alternative implies that either \( \Pi(z) \notin L^2 \) for all \( z \) \( (\text{im} z \neq 0) \) (the point case) or \( \Pi(z) \in L^2 \) for all such \( z \) (the disk case). The criterion for definiteness of the moment problem is thereby reformulated in terms of the spectral theory of Jacobian matrices.

---

16 The latter follows from the positivity of the leading coefficients.

17 Putting this differently we can say that equation (39) (with \( x \) replaced by \( z \) \( (\text{im} z \neq 0) \)) either has a unique solution up to normalisation in \( L^2 \) or all its solutions belong to \( L^2 \) and moreover for all \( z \) the same case results.
2.8. The Trigonometric Moment Problem. Suppose there is given on the circumference of the unit circle a measure \( \mu \) which is generated by a non-decreasing function \( \sigma(x) \) \( (-\pi \leq x \leq \pi) \). Then the numbers

\[
c_n = \int_{|z|=1} z^n \, d\mu = \int_{-\pi}^{\pi} e^{inx} \, d\sigma(x) \quad (n = 0, \pm 1, \pm 2, \ldots)
\]

are called the power moments of the measure \( \mu \) or the trigonometric moments of the function \( \sigma \). Here again the problem is to give an intrinsic characterisation of the class of moment sequences \( (c_n)_{n=0}^{\infty} \). Analogous to the theorem of Hamburger we have

Theorem of F. Riesz and Herglotz. In order that \( (c_n)_{n=0}^{\infty} \) be a moment sequence, it is necessary and sufficient that the Toeplitz matrix \( (c_{k-j})_{j=0}^{\infty} \) be positive, i.e.

\[
\sum_{k,j=0}^{n} c_{k-j} \xi_k \xi_j \geq 0 \quad (41)
\]

for all \( n = 0, 1, 2, \ldots \) and for all \( (\xi_1, \ldots, \xi_n) \in \mathbb{C}^n \).

The corresponding measure \( \mu \) is unique.

The necessity is trivial. Sufficiency: with \( \xi_k = e^{ikt} \) \( (0 \leq k \leq n; t \in \mathbb{R}) \) we have from (41)

\[
\sum_{|m| \leq n} (n + 1 - |m|) c_m e^{imt} \geq 0
\]

(see (29)). Therefore the functions

\[
\sigma_n(x) = \int_{-\pi}^{\pi} \sum_{|m| \leq n} \left(1 - \frac{|m|}{n + 1}\right) c_m e^{imt} \, dt
\]

are non-decreasing. Moreover they are uniformly bounded: \( \sigma_n(\pi) = 2\pi c_0 \). By a well-known theorem of Helly there exists a non-decreasing function \( \sigma(x) \) and a subsequence \( (\sigma_{n_k}(x)) \) which converges weakly to \( \sigma(x) \) in the sense that

\[
\lim_{k \to \infty} \int_{-\pi}^{\pi} f(x) \, d\sigma_{n_k}(x) = \int_{-\pi}^{\pi} f(x) \, d\sigma(x)
\]

for all \( f \in C[-\pi, \pi] \). The function \( \sigma(x) \) is the desired function.

The uniqueness of the desired measure is a simple consequence of the uniqueness of the function with given Fourier coefficients. \( \Box \)

The trigonometric moment problem has a natural continuous analogue. It is to give an intrinsic description of the class of functions \( c(\lambda) \) \( (-\infty < \lambda < \infty) \) admitting a representation of the form

\[
c(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} \, d\sigma(x), \quad (42)
\]

where \( \sigma(x) \) is a non-decreasing function. A necessary and sufficient condition for this, established by Bochner (1932) and independently by A.Ya. Khinchin (1937),
Chapter 1. Classical Concrete Problems

is the analogue of the Riesz-Herglotz theorem. It consists of continuity and positivity of the function \( c(\lambda) \). The latter means that

\[
\sum_{k,j=1}^{n} c(\lambda_k - \lambda_j) \xi_k \overline{\xi_j} \geq 0
\]

for all \((\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n\) and all \((\xi_1, \ldots, \xi_n) \in \mathbb{C}^n\) \((n = 1, 2, 3, \ldots)\). In the terminology of probability theory we can say that the characteristic functions of the probability laws are exactly the continuous positive functions satisfying the normalisation condition \( c(0) = 1 \).

2.9. The Fourier Integral. This continuous analogue of Fourier series appeared (1822) in the solution of the problem of heat conduction in an infinite rod. In this case the boundary conditions drop out and only the initial condition \( u(x, 0) = \phi(x) \) remains. The formal solution obtained by separation of variables has the form (for \( \kappa = 1 \))

\[
u(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\lambda) e^{-2\pi i \lambda x} d\lambda,
\]

where

\[
E(\lambda) = \int_{-\infty}^{\infty} \phi(x) e^{-i\lambda x} dx.
\]

Here it is required that \( \phi \in L^1 \) although, after eliminating \( E(\lambda) \) by means of the heuristic transformations from (43) and (44), the formula

\[
u(x, t) = \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/(4t)} \phi(y) dy
\]

is obtained, which applies to a much wider class of initial functions \( \phi \). Thus in this instance the Fourier transform \( E(\lambda) \) played an important but temporary role. This particular circumstance cannot detract from the significance of the Fourier transform and its generalisations both as a tool for investigating numerous problems and as an independent object.

In many ways the theory of the Fourier integral parallels that of Fourier series. The formal inversion of the Fourier transform has the form

\[
\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(\lambda) e^{2\pi i \lambda x} d\lambda,
\]

but in fact it is valid under very stringent restrictions, for example, if \( E(\lambda) \) is

\[18\] This is clear from (43) for \( t = 0 \). Formally combining (44) and (44*) we obtain the orthogonality relation

\[
\int_{-\infty}^{\infty} e^{i\lambda x} e^{-i\lambda y} d\lambda = 2\pi \delta(x - y),
\]

which is the typical form for the continuous situation.
The inversion formula is such that
\[ \int_0^\infty \phi(\xi) \, d\xi = \lim_{N \to \infty} \frac{1}{2\pi} \int_{-N}^N E(\lambda) \frac{e^{i\lambda x} - 1}{i\lambda} \, d\lambda \]
(from which \( \phi \) may be determined almost everywhere by differentiation). Consequently \( \phi \) is uniquely determined by its Fourier transform.

The \( L^2 \)-theory of the Fourier integral is contained in the following theorem of Plancherel. Let \( \phi \) belong to \( L^2(-\infty, \infty) \). Then the Fourier-Plancherel transform
\[ E(\lambda) = \lim_{N \to \infty} \frac{1}{\sqrt{(2\pi)}} \int_{-N}^N \phi(x) e^{-i\lambda x} \, dx \]
exists in mean square and the inversion formula
\[ \phi(x) = \lim_{N \to \infty} \frac{1}{\sqrt{(2\pi)}} \int_{-N}^N E(\lambda) e^{i\lambda x} \, d\lambda \]
holds (convergence in mean square) as does the corresponding Parseval equality
\[ \int_{-\infty}^\infty |E(\lambda)|^2 \, d\lambda = \int_{-\infty}^\infty |\phi(x)|^2 \, dx. \]

Hence it is clear that the Fourier \( L^2 \)-transform is an isometric isomorphism of the space \( L^2(-\infty, \infty) \) of functions in \( x \) and in \( \lambda \). These two spaces are identified formally, but such a point of view is not always acceptable (for example, in quantum mechanics the variable \( x \) may have the sense of a coordinate and then the variable \( \lambda \) is associated with an impulse of the particle; their physical identification contradicts the principle of uncertainty).

We can also refer to the formula (42) as a variant of the Fourier transform: this is the Fourier-Stieltjes transform of the function \( \sigma(x) \) or the Fourier transform of the measure \( \mu_\sigma \). We note that the Bochner-Khinchin theorem can be utilised in the construction of the \( L^2 \)-theory of the Fourier transform.

Let us consider a function \( \phi(x) \) which is zero for \( x < 0 \). If this \( \phi \in L^2(0, \infty) \) then its Fourier-Plancherel transform has the form
\[ E(\lambda) = \lim_{N \to \infty} \frac{1}{\sqrt{(2\pi)}} \int_0^N \phi(x) e^{-i\lambda x} \, dx. \]
The specific character of this case lies in the fact that the values of \( E(\lambda) \) are almost everywhere boundary values\(^{21}\) for the function
\[ \hat{E}(\lambda) = \frac{1}{\sqrt{(2\pi)}} \int_0^\infty \phi(x) e^{-i\lambda x} \, dx \]
which is analytic on the half-plane \( \text{im } \lambda < 0 \). Moreover

\(^{19}\)In the general case \( E(\lambda) \) is continuous, tends to zero as \( |\lambda| \to \infty \), but may not belong to \( L^1(-\infty, \infty) \).
\(^{20}\)Or Fourier \( L^2 \)-transform.
\(^{21}\)Along non-tangential paths.
\[
\int_{-\infty}^{\infty} |\hat{E}(\alpha - \beta i)|^2 \, d\alpha = \int_{0}^{\infty} |\phi(x)|^2 e^{-2\beta x} \, dx \leq \int_{0}^{\infty} |\phi(x)|^2 \, dx,
\]
i.e. \(\hat{E}\) belongs to the linear space \(H^2\) of functions \(f(\lambda)\) which are analytic in the half-plane \(\text{Im} \lambda < 0\) and satisfy the condition
\[
\sup_{\beta > 0} \int_{-\infty}^{\infty} |f(\alpha - \beta i)|^2 \, d\alpha < \infty.
\]
The analogous definition for \(p \geq 1\) gives a whole family of spaces, the Hardy classes \(H^p\).

The theorems stated below on functions of class \(H^2\) are due to Wiener and Paley.

**Theorem 1.** Any function \(f \in H^2\) can be represented in the form
\[
f(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \phi(x) e^{-i\lambda x} \, dx \quad (\text{Im} \lambda < 0),
\]
where \(\phi \in L^2(0, \infty)\).

The function \(\phi\) in (45) is obtained as the inverse Fourier transform of the boundary function \(f_0\) for \(f\), taken for \(x > 0\) (for \(x < 0\) this procedure gives zero). The existence almost everywhere of the boundary values for \(f \in H^2\) (and in general \(H^p\)) and the fact that \(f_0\) belongs to \(L^2\) (respectively \(L^p\)) follow from the classical theory of Fatou.

The boundary function \(f_0\) for \(f \in H^2\) \((f \neq 0)\) has the important property
\[
\int_{-\infty}^{\infty} \frac{|\ln |f_0(\lambda)||}{1 + \lambda^2} \, d\lambda < \infty
\]
and, conversely, for any function \(\psi \in L^2(-\infty, \infty), \psi \geq 0\) such that
\[
\int_{-\infty}^{\infty} \frac{|\ln \psi(\lambda)||}{1 + \lambda^2} \, d\lambda < \infty,
\]
there exists \(f \in H^2\) for which \(|f_0| = \psi\).

A function \(\phi(x)\) on the whole axis (or on the semiaxis) is said to be of compact support if it vanishes outside of some finite interval \((a, b)\). If such a function \(\phi \in L^2(a, b)\) then its Fourier-Plancherel transform reduces to
\[
E(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} \phi(x) e^{-i\lambda x} \, dx.
\]
By means of this identity \(E(\lambda)\) is extended to the whole complex plane as an entire function of exponential type \(\rho \leq \max(|a|, |b|)\) and in addition, by the theorem of Plancherel, it belongs to \(L^2\) on the real axis.

**Theorem 2.** If \(f(\lambda)\) is an entire function of exponential type \(\rho\) which belongs to \(L^2\) on the real axis, then it can be represented in the form
where $\phi \in L^2(-\rho, \rho)$.

The class of entire functions described in this theorem is denoted by $B^2_p$.

We note further a continuous analogue of the theorem on functions of class $W$ which is important for applications.

**Theorem of Wiener.** If $\phi \in L^1(-\infty, \infty)$ and its Fourier transform $E(\lambda) \neq 1$ $(-\infty < \lambda < \infty)$ then the function $(1 - E(\lambda))^{-1} - 1$ is the Fourier transform of a function $\psi \in L^1(-\infty, \infty)$.

$L^1(-\infty, \infty)$ is a Banach algebra (without identity) under convolution

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t - \tau)g(\tau) \, d\tau$$

as multiplication. The Fourier transform transforms it into the *Wiener algebra* $\mathcal{W}(-\infty, \infty)$ of functions with the usual (pointwise) multiplication.

**2.10. The Laplace Transform.** We shall say that the function $f(t)$ $(t \geq 0)$ belongs to class $A$ if it is Lebesgue measurable and there exists $k = k(f)$ such that $f(t) = O(e^{kt})$ almost everywhere. The class $A$ is a linear space. If $f \in A$ then the analytic function

$$\tilde{f}(\lambda) = \int_0^{\infty} f(t)e^{-\lambda t} \, dt$$

is defined in the half-plane $\Re \lambda > k(f)$. The class $\tilde{A}$ of all such functions is also linear. The homomorphism $f \mapsto \tilde{f}$ is called the *Laplace transform* and under it $f$ is called the preimage for its image $\tilde{f}$. The preimage is uniquely determined for a given image, and moreover the Riemann-Mellin inversion formula holds:

$$\int_0^{t} f(\tau) \, d\tau = \lim_{\beta \to -\infty} \frac{1}{2\pi i} \int_{\alpha - \beta i}^{\alpha + \beta i} \frac{\tilde{f}(\lambda)}{\lambda} e^{\lambda t} \, d\lambda \quad (\alpha > k(f)).$$

This formula opens up a wide range of possibilities for calculations using the theory of residues. For example, let $\tilde{f}(\lambda)$ be continued to the whole plane as a meromorphic function with poles $\{\lambda_k\}_{1}^{\infty}$. To make matters precise let us assume $k(f) \geq 0$ and put $\lambda_0 = 0$. If $\tilde{f}$ is bounded on some sequence of left semicircles with diameters $[\alpha - \beta_n i, \alpha + \beta_n i]$ ($\beta_n \to \infty$), then

$$\int_0^{t} f(\tau) \, d\tau = \lim_{n \to \infty} \sum_{|\lambda - \alpha| < \beta_n} \text{Res} \left[ \frac{\tilde{f}(\lambda)}{\lambda} e^{\lambda t} \right],$$

the limit being uniform on each finite interval. The residues in (46) have the form $P_k(t)e^{\lambda_k t}$, where $P_k$ is a polynomial of degree $\leq n_k - 1$ and $n_k$ is the order of the pole $\lambda_k$ for $\frac{\tilde{f}(\lambda)}{\lambda}$. Consequently, under the stated conditions the integral of $f$ (and
under certain additional requirements also \( f \) itself) turns out to be the sum with brackets of the series of quasipolynomials \( \sum_k P_k(t)e^{\lambda_k t} \) (here the \( \lambda_k \) are indexed according to increasing \( |\lambda_k - \alpha| \)); moreover convergence is uniform on each finite interval.

Formula (46) is called the Expansion Theorem.

We note further that if the function \( f \in A \) is continuous then

\[
\lim_{n \to \infty} \left( \frac{(-1)^n}{n!} \frac{d}{dt} \left( \frac{n}{t} \right) \right)^{n+1} f(t) = \lim_{n \to \infty} \frac{1}{n!} \frac{d}{dt} \left( \frac{n}{t} \right)^n f(t),
\]

the limit being uniform on each finite interval \((a, b) (a > 0)\). The specific character of this formula of Widder, which goes back for its idea to Laplace’s 'method of larger numbers', lies in the fact that use is made in it of the value of the Laplace transform only for large \( \lambda > 0 \).

A most important property of the Laplace transform is that it transforms the operation of differentiation in \( A \) under the condition \( f(0) = 0 \) into the operation of multiplication by \( \lambda \) in \( \mathbb{L} \). The more general relation

\[
\hat{f}'(\lambda) = \lambda \hat{f}(\lambda) - f(0)
\]

is valid for all absolutely continuous \( f \) such that \( f' \in A \) (in this case \( f \in A \) automatically).

If \( f, g \) are two functions of class \( A \) then the function

\[
\int_0^t f(\tau) g(t - \tau) d\tau
\]

is called their Laplace convolution and is denoted as before by \( f \ast g \). It is easy to see that it also belongs to the class \( A \) and \( \hat{f} \ast \hat{g} = \hat{f} \hat{g} \), i.e. the Laplace transform transforms convolution into multiplication. We can say that the Laplace transform (as also the Fourier transform) expresses algebraically basic operations of classical analysis. On this is founded the most commonly used modern version of operational calculus, which was formally established by Heaviside in the last century as a tool for the solution of differential equations and certain other forms of functional equations. We give several examples on this theme, first of all the traditional

**Example 1.** Let us consider the linear differential equation of \( m \)th order

\[
D[y] = y^{(m)} + a_1 y^{(m-1)} + \cdots + a_m y = f(t)
\]

with constant coefficients and right hand side \( f \in A \). Let \( y \) be the solution under
the initial conditions \( y^{(k)}(0) = 0 (0 \leq k \leq m - 1) \). It is easy to see that \( y \in A \). From (47) we have

\[
y(\lambda) = \frac{\tilde{f}(\lambda)}{D(\lambda)},
\]

where \( D(\lambda) = \lambda^m + a_1 \lambda^{m-1} + \cdots + a_m \) is the characteristic polynomial. The rational function \( \frac{1}{D(\lambda)} \) is the Laplace transform of the quasipolynomial

\[
Q(t) = \sum \text{Res}_{\lambda=\lambda_n} \left[ \frac{e^{\lambda t}}{D(\lambda)} \right],
\]

where the sum is taken over all roots of the polynomial \( D(\lambda) \). According to the convolution theorem

\[
Q \ast f(\lambda) = \frac{\tilde{f}(\lambda)}{D(\lambda)}
\]

and because of the uniqueness of the preimage, \( Q \ast f = y \), i.e.\(^{24}\)

\[
y(t) = \int_0^t Q(t - \tau)f(\tau) \, d\tau.
\]

The quasipolynomial \( Q \) is called the Cauchy function of the operator \( D \).

**Example 2.** Now let the right hand side in equation (47) be one-periodic: \( f(t + 1) = f(t) \). Let us assume that the characteristic polynomial does not have roots of the form \( in \) (where \( n \) is an integer). Then there is a unique one-periodic solution, namely

\[
y(t) = \sum_{n=-\infty}^{\infty} \frac{c_n}{D(it)} e^{int} \quad (48)
\]

(the \( c_n \) are the Fourier coefficients of the function \( f \)). We shall obtain this solution by applying the Laplace transform:

\[
D(\lambda)\tilde{y}(\lambda) + \Delta(\lambda) = \tilde{f}(\lambda).
\]

Here \( \Delta(\lambda) \) is a polynomial of degree \( \leq m - 1 \) whose coefficients are linear combinations of the initial values \( y(0), \ldots, y^{(m-1)}(0) \). From the periodicity of \( f \) it follows that

\[
\tilde{f}(\lambda) = \frac{f_1(\lambda)}{e^{\lambda} - 1}, \quad f_1(\lambda) = \int_0^1 f(1 - t)e^{\lambda t} \, dt
\]

and \( \tilde{y}(\lambda) \) is expressed in a similar way. Therefore from (49)

\[
\frac{y_1(\lambda)}{e^{\lambda} - 1} = \frac{f_1(\lambda)}{D(\lambda)(e^{\lambda} - 1)} - \frac{\Delta(\lambda)}{D(\lambda)}.
\]

\(^{24}\) This final result is valid for any locally summable \( f \).
Since the left hand side is regular at the root $\lambda_k$ of the polynomial $D(\lambda)$,
\[
\text{Res}_{\lambda = \lambda_k} \left[ \frac{\Delta(\lambda)}{D(\lambda)} e^{\lambda t} \right] = \text{Res}_{\lambda = \lambda_k} \left[ \frac{f_1(\lambda)e^{\lambda t}}{D(\lambda)(e^{\lambda} - 1)} \right].
\]
Thus (49) leads to the solution
\[
y(t) = \int_0^t Q(t - \tau)f(\tau) \, d\tau - \sum_{\lambda = \lambda_k} \text{Res}_{\lambda = \lambda_k} \left[ \frac{f_1(\lambda)e^{\lambda t}}{D(\lambda)(e^{\lambda} - 1)} \right]
\]
in finite form, in contrast to formula (48).

**Example 3.** Let $0 < \alpha < 1$. The function
\[
f_\alpha(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t - \tau)^{\alpha-1}f(\tau) \, d\tau
\]
is called the integral of order $\alpha$ of the function $f \in A$ (i.e. the convolution of the functions $\frac{1}{\Gamma(\alpha)} t^{\alpha-1}$ and $f$). If $f_\alpha = g$ is given then (50) can be regarded as an equation for $f$ (Abel integral equation). By the convolution theorem
\[
f_\alpha(\lambda) = \lambda^\alpha \hat{g}(\lambda) = \lambda \hat{g}(1 - \alpha)(\lambda),
\]
from which we obtain
\[
f(t) = \frac{d}{dt} \left( \frac{1}{\Gamma(1 - \alpha)} \int_0^t (t - \tau)^{-\alpha}g(\tau) \, d\tau \right).
\]
The operator (51) is naturally called the derivative of order $\alpha$.

The Laplace transform can be applied not only to functions but also to measures:
\[
S(\lambda) = \int_0^\infty e^{\lambda t} \, d\sigma(t) \quad (\text{re } \lambda > k).
\]
Here $\sigma$ is a function of bounded variation on each interval $[0, T]$ ($T > 0$) subject to the requirement
\[
\sigma(t) = O(e^{-kt}).
\]
In connection with such functions $\sigma$ we use the term Laplace-Stieltjes transform.

This scheme covers in particular Dirichlet series
\[
\sum_{n=1}^{\infty} \frac{a_n}{n^\lambda} \quad (\text{re } \lambda > k)
\]
with arbitrary coefficients $a_n$ satisfying the condition
\[
\left| \sum_{n=1}^N a_n \right| = O(N^k).
\]
When $a_n = 1$ the series (53) defines the Riemann $\zeta$-function which plays an
extremely important role in analytic number theory, where it and certain other Dirichlet series find application.\textsuperscript{25}

If the function $\sigma$ is non-decreasing and satisfies the condition (52) for every $k > 0$ and $S(\lambda) (\lambda > 0)$ is its Laplace-Stieltjes transform then the function $S(-\lambda)$ is absolutely monotonic, i.e. it is infinitely differentiable and all its derivatives (including itself) are non-negative.

**Theorem of Bernshtein.** Any absolutely monotonic function $f(x) (x < 0)$ may be represented in the form

$$f(x) = \int_0^\infty e^{xt} d\sigma(t),$$

where $\sigma$ is a non-decreasing function.

This result can be regarded as a solution of the exponential moment problem.

In contrast to the Fourier transform, the Laplace transform is applicable to increasing functions but it is nevertheless required that their growth is no greater than exponential. However it is possible to modify the Laplace transform in such a way that the restrictions connected with growth rate are completely removed and moreover it becomes possible to work with functions on finite intervals in the same way as on infinite intervals. In fact, let $0 < b \leq \infty$ and suppose that the function $f(t) (0 < t < b)$ is locally summable. The local Laplace transform (l.L.t.) of the function $f$ is that function $F(\lambda)$, which is defined and locally bounded for sufficiently large $\lambda > 0$, such that for all $\beta \in (0, b)$

$$F(\lambda) = \int_0^\mu f(t)e^{-\lambda t} dt + \varepsilon(\lambda, \beta),$$

where the remainder $\varepsilon(\lambda, \beta)$ satisfies the condition

$$\lim_{\lambda \to +\infty} \frac{\ln|\varepsilon(\lambda, \beta)|}{\lambda} \leq -\beta,$$

i.e. for $\lambda \to +\infty$ it has exponential type $\leq -\beta$. If $f \in A$ then its Laplace transform is its l.L.t. We note that if the l.L.t. exists for a given preimage $f$ then it is defined up to a term of exponential type $\leq -b$ as $\lambda \to +\infty$. The question of conditions for the existence of the l.L.t. seems to be non-trivial but in applications it usually has an automatic positive resolution. The uniqueness of the determining function for a given l.L.t. is an essential fact. This results from the following procedure for inverting the l.L.t. (Yu.I. Lyubich, 1966).

**Theorem.** Let $F(\lambda) (\lambda \geq \mu > 0)$ be the l.L.t. of $f(t) (0 < t < b)$. Then the function

$$\Phi(\zeta) = \frac{1}{2\pi i} \int_\mu^\infty \frac{F(\lambda)}{\lambda} e^{\lambda \zeta} d\lambda \quad (\text{re} \zeta < 0)$$

Dirichlet introduced these series for the proof of his theorem (1837) that there are infinitely many primes in any arithmetic progression $(a + nd)_{n=0}^\infty$ where $a, d$ are relatively prime natural numbers.
can be continued analytically to the domain

\[ P_b = \{ \zeta: \text{re} \zeta < b, \zeta \notin [0, b) \} \]

and

\[ \lim_{s \downarrow 0} \{ \Phi(t + is) - \Phi(t - is) \} = \int_0^t \Phi'(\tau) \, d\tau \quad (0 < t < b). \]  \tag{55} \]

If \( f \) is smooth then the function

\[ \Psi(\zeta) = \frac{1}{2\pi i} \int_{\mu}^{\infty} F(\lambda) e^{\lambda \zeta} \, d\lambda \quad (\text{re} \zeta < 0) \]

can be continued analytically to the domain \( P_b \), it has limiting values \( \Psi(t \pm i0) \) and

\[ \Psi(t + i0) - \Psi(t - i0) = f(t) \quad (0 < t < b). \]

In the series of applications it is necessary to weaken the requirement (54) by allowing \( \lambda \) to tend to infinity only through some set \( M \) of complete relative measure, i.e. the measure of the intersection \( M \cap [0, r] \) is asymptotically equal to \( r (r \to +\infty) \). This modification (which we shall denote by l.L.t.) is adequate if the l.L.t. \( F(\lambda) \) can be extended to a function which is analytic on some angular domain \( -\alpha < \text{arg} \lambda < \beta \left( 0 < \alpha, \beta < \frac{\pi}{2} \right) \) and increases no faster than exponentially. In this case the inversion formula (55) and, consequently, the uniqueness theorem for the preimage remain valid (Yu.I. Lyubich and V.A. Tkachenko, 1965). Hence we obtain the following analogue of the Wiener-Paley theorem: if the l.L.t. \( F(\lambda) \) is an entire function of exponential type \( \sigma \) then it is a finite Laplace transform:

\[ F(\lambda) = F(\lambda; \sigma) \equiv \int_0^\sigma f(t)e^{-\lambda t} \, dt. \]

In particular, if \( \sigma = 0 \) then \( F = 0 \).

Let us consider one of the applications of the local Laplace transform.

According to Delsarte (1934) a continuous function \( f(t) \) \( (-\infty < t < \infty) \) is said to be periodic in mean (p.im.) if for some finite \( a, b \) \( (a < b) \) it satisfies an integral equation of the form

\[ \int_a^b f(t + s) \, d\omega(s) = 0 \quad (-\infty < t < \infty) \]

where \( \omega \) is a complex function of bounded variation \( (\omega \neq \text{constant} \text{ on neighbourhoods of the points} \ a, b) \). In the case \( \omega(s) = s \) the equation is equivalent to the periodicity (with period \( b - a \) of the function \( f \) and its mean over the period being zero. Naturally we expect to have for p.im. a theory similar to that of Fourier series. In fact from the equation for \( f \) we may formally calculate the Laplace transform:

\[ \tilde{f}(\lambda) = \frac{\Phi_f(\lambda)}{\Delta(\lambda)}, \]
where
\[ \Delta(\lambda) = \int_a^b e^{-\lambda s} \, ds, \quad \Phi_f(\lambda) = \int_a^b e^{\lambda s} \, ds \int_a^s f(t) e^{-\mu t} \, dt. \]

The function \( \tilde{f} \) is meromorphic and the roots \( \lambda_1, \lambda_2, \lambda_3, \ldots \) of the characteristic equation \( \Delta(\lambda) = 0 \) are its poles. Formally applying the Expansion Theorem (in differentiated form) we arrive at the generalised Fourier series:
\[ f(t) \sim \sum_{n=1}^{\infty} P_n(t) e^{\lambda_n t}, \]
where
\[ P_n(t) = e^{-\lambda_n t} \operatorname{Res}_{\lambda=\lambda_n} \left[ \tilde{f}(\lambda) e^{\lambda t} \right] \]
is a polynomial of degree \( \leq \nu_n - 1, \nu_n \) being the multiplicity of the root \( \lambda_n \). It turns out that this series is uniformly summable to \( f \) on each finite interval by a certain special method (L. Schwartz, 1947). From this follows the uniqueness theorem: if the generalised Fourier series of the p.i.m. function \( f \) is zero then \( f = 0 \). However, it is possible to establish this fact independently by using the l.L.t.' if we note that \( \tilde{f}(\lambda) \) is such. In the case of a zero series \( \tilde{f} \) is an entire function of exponential type \( \sigma = 0 \) and consequently \( \tilde{f} = 0 \). We emphasise that hidden in this argument is a non-trivial application of the theory of growth of entire functions.

If the p.i.m. function \( f \) is equal to zero on \([a, b]\) then it is equal to zero on the whole axis because \( \Phi_f = 0 \) and hence \( \tilde{f} = 0 \). Also it is easy to deduce this theorem on the unique determination of a p.i.m. function by its values on the period' from Titchmarsh’s convolution theorem: if the functions \( f(t), g(t) \) \((0 < t < b)\) are locally summable on \((0, b)\) and if the Laplace convolution \( f \ast g \) is equal to zero while \( f \big|_{(0, \varepsilon)} \neq 0 \) for arbitrarily small \( \varepsilon > 0 \), then \( g = 0 \). However Titchmarsh’s theorem can also be established by means of the local Laplace transform. In fact the finite Laplace transforms \( F(\lambda; \beta), G(\lambda; \beta) \) \((0 < \lambda < \beta)\) are connected by the relation
\[ F(\lambda; \beta)G(\lambda; \beta) - R(\lambda; \beta) = 0 \]
where
\[ R(\lambda; \beta) = \int_0^\beta g(t) \delta(t) e^{-\lambda t} \, dt \int_{\beta-t}^\beta f(s) e^{-\lambda s} \, ds = O(e^{-\beta \lambda}). \]
But then the l.L.t.' for \( g \) is equal to zero.

### 2.11. The Sturm-Liouville Problem

This designation relates to the problem of determining the eigenvalues and eigenfunctions of the linear differential operator
\[ L[y] = \frac{1}{r(x)} \left[ - \frac{d}{dx} \left( p(x) \frac{dy}{dx} \right) + q(x)y \right] \quad (a \leq x \leq b) \]
\((r(x), q(x) \) are continuous functions, \( p(x) \) is smooth, \( r(x) > 0, p(x) > 0, q(x) \) is real\) with boundary conditions
Chapter 1. Classical Concrete Problems

\[ y'(a) - h_1 y(a) = 0, \quad y'(b) + h_2 y(b) = 0 \]

\((h_1, h_2 \in \mathbb{R} \cup \{\infty\});\) if, for example, \(h_1 = \infty\) then the corresponding condition has the form \(y(a) = 0\). In 1836 Sturm and Liouville independently constructed a systematic theory of this problem (with \(q(x) \equiv 0\)), which generalises the theory of Fourier series of the form (21) at least for smooth functions. In the 20th century the Sturm-Liouville problem and its generalisations became an essential attribute of the mathematical apparatus of quantum mechanics. The point is that in quantum mechanics physical quantities are described by linear operators\(^{26}\), their possible values are the eigenvalues of the corresponding operators and they are realised in the states of the quantum mechanical system, these being defined by the corresponding eigenfunctions. In particular \(-y'' + q(x)y\) is the operator of total energy\(^{27}\) (Hamiltonian\(^{28}\)): \((-y'')\) corresponds to kinetic energy, \(q(x)y\) to potential energy. The eigenvalues of an energy operator are the energy levels of the system.

We shall consider the operator \(L\) on \(L^2(a, b)\), \((\sigma(x) = \int_a^b r(\xi)d\xi)\), assuming for simplicity that its domain of definition \(D(L)\) consists of those functions of class \(C^2[a, b]\) which satisfy the boundary conditions. A basic property of the operator \(L\) is contained in the identity

\[ (L[y_1], y_2) = (y_1, L[y_2]) \quad (y_1, y_2 \in D(L)), \]

which allows us to call \(L\) symmetric by analogy with the finite-dimensional case. As a consequence of this, the eigenvalues of the operator are real and eigenfunctions corresponding to a pair of distinct eigenvalues are orthogonal. Consequently the set of eigenvalues is no more than countable. Each eigenspace is one-dimensional (it cannot be two-dimensional because of the presence of the boundary conditions).

**Theorem.** The set of eigenvalues of the Sturm-Liouville problem can be represented as a sequence \(\lambda_1 < \lambda_2 < \cdots < \lambda_n < \cdots\) tending to \(\infty\). Any function \(f \in D(L)\) can be expanded in a uniformly convergent generalised Fourier series with respect to the orthonormalised system of eigenfunctions:

\[
 f(x) = \sum_{n=1}^{\infty} c_n y_n(x) \quad \left( c_n = (f, y_n) = \int_a^b f(x)y_n(x)r(x) \, dx \right).
\]

For any \(f \in L^2(a, b)\) the generalised Fourier series converges to \(f\) in \(L^2(a, b)\), which is equivalent to Parseval's equality

\[
 \sum_{n=1}^{\infty} |c_n|^2 = \int_a^b |f(x)|^2 r(x) \, dx
\]

being satisfied.

\(^{26}\)Initially without linearisation!

\(^{27}\)For the given boundary conditions.

\(^{28}\)By analogy with the Hamiltonian function in classical mechanics. We note that the operator is also called the Schrödinger operator.
The proof of this theorem was outlined in essence in the works of Sturm and Liouville but it was not until 1896 that a rigorous version was given by V.A. Steklov.

One of the methods of proof is based on the variational theory of eigenvalues developed by Fischer and Courant. In this theory the equation \( L[y] = \lambda y \) is regarded as an Euler-Lagrange equation for the problem of finding the minimum of the quadratic functional \( (L[y], y) \) under the constraint condition \( (y, y) = 1 \). It turns out that the desired minimum is the least eigenvalue \( \lambda_1 \) and is attained at a corresponding eigenfunction \( y_1 \). If we already have the eigenvalues \( \lambda_1 < \cdots < \lambda_n \) and corresponding eigenfunctions \( y_1, \ldots, y_n \) then

\[
\lambda_{n+1} = \min\{ (L[y], y) : (y, y) = 1, (y, y_k) = 0 \ (1 \leq k \leq n) \}
\]

and this minimum is attained at a corresponding eigenfunction \( y_{n+1} \). Thus we obtain an infinite increasing sequence \( (\lambda_n) \) and it is easy to prove that \( \lambda_n \to \infty \). The inequality

\[
\left\| f - \sum_{k=1}^{n} c_k y_k \right\| \leq \frac{1}{\lambda_{n+1}} \| L[f] \| \quad (f \in D(L))
\]

follows from the construction just described and letting \( n \to \infty \) we obtain from it the expansion in \( L^2(a, b) \):

\[
f = \sum_{k=1}^{\infty} c_k y_k. \tag{56}
\]

The condition \( f \in D(L) \) is not essential for convergence in \( L^2 \). If nevertheless \( f \not\in D(L) \), we therefore have in \( L^2 \)

\[
L[f] = \sum_{k=1}^{\infty} (L[f], y_k) y_k = \sum_{k=1}^{\infty} c_k \lambda_k y_k.
\]

The uniform convergence of the series (56) follows from this as a result of the bound

\[
|f(x)| \leq \text{const} \| L[f] \|
\]

(requiring a separate proof).

2.12. The Schrödinger Operator on the Semiaxis. Let us consider the operator \( L[y] = -y'' + q(x)y \) \((0 \leq y < \infty)\) with boundary condition \( y'(0) - hy(0) = 0 \) \((h \in \mathbb{R} \cup \{\infty\})\). We assume the potential \( q(x) \) to be real and (for simplicity) continuous. On \( L^2(0, \infty) \) the set of eigenvalues of such an operator can turn out to be finite or even empty. For example, with \( q(x) \equiv 0, h = 0 \) we require non-trivial solutions of the problem \( y'' + \lambda y = 0, y'(0) = 0 \), these being (arbitrary multiples of) \( \psi(x, \lambda) = \cos \sqrt{\lambda}x \). None of these belongs to \( L^2(0, \infty) \). We note however that \( \psi(x, \lambda) \) is bounded for each \( \lambda \geq 0 \) which suggests weakening the requirements as \( x \to \infty \) on the eigenfunctions, i.e. to let them lie outside of the class \( L^2 \). From the point of view of quantum mechanics this is admissible since
the energy spectrum can be both discrete and continuous. Unfortunately, however, a requirement such as boundedness is not sufficient in general (V.P. Maslov, 1961). The problem arising here is resolved in a certain sense only in a sufficiently wide context of functional analysis (I.M. Gel'fand – A.G. Kostyu-chenko, 1955). It is possible to by-pass the difficulty by analytical means if we introduce a particular collection of 'generalised' eigenfunctions for which we do not formulate explicit boundedness of growth or decrease at infinity. This approach, which was opened up by the pioneering research of H. Weyl, was developed by Titchmarsh (1946) and then substantially simplified by B.M. Levitan (1956).

If \( h = \tan \alpha \left( \frac{-\pi}{2} < \alpha \leq \frac{\pi}{2} \right) \), the boundary condition at zero can be expressed in the form \( y'(0) \cos \alpha - y(0) \sin \alpha = 0 \). Let us denote by \( \psi(x, \lambda) \) the solution of the equation \( L[y] = \lambda y \) which satisfies the boundary condition through the initial conditions \( \psi(0, \lambda) = \cos \alpha, \psi'(0, \lambda) = \sin \alpha \). We take an arbitrary interval \([0, b]\) and consider the Sturm-Liouville problem \( L[y] = \lambda y \) with the same condition at zero and any desired condition \( y'(b) \cos \beta + y(b) \sin \beta = 0 \). If \( \{\lambda_n\}^\infty_1 \) is the set of its eigenvalues then \( \{\psi(x, \lambda_n)\}^\infty_1 \) is a set of eigenfunctions; in general, however, these are not normalised. Put \( \alpha_n = \|\psi(\cdot, \lambda)\|_{L^2(0, b)} \). Then for any \( f \in L^2(0, b) \) we will have in mean square

\[
|f(x)|^2 = \sum_{n=1}^\infty \frac{\psi(x, \lambda_n) \psi(x, \lambda_n)}{\alpha_n^2}.
\]

Let us introduce the non-decreasing function

\[
\rho_b(\lambda) = \sum_{\lambda_n < \lambda} \frac{1}{\alpha_n^2}.
\]

Then

\[
f(x) = \int_{-\infty}^\infty F_b(\lambda) \psi(x, \lambda) \, d\rho_b(\lambda),
\]

where

\[
F_b(\lambda) = \int_0^b f(x) \psi(x, \lambda) \, dx.
\]

Parseval equality in similar notation has the form

\[
\|f\|^2_{L^2(0, b)} = \sum_{n=1}^\infty \frac{\psi(x, \lambda_n) \psi(x, \lambda_n)}{\alpha_n^2}.
\]

In general it is mixed. For example, in the spectrum of the hydrogen atom there is a countable system of discrete energy levels \( E_n < 0 \) corresponding to the states of an electron which is connected with the nucleus and a continuous spectrum \( E \geq 0 \) corresponding to the possibility of unbounded escape from the nucleus.

The first work of H. Weyl on the Schrödinger operator (so called much later) on the semiaxis appeared in 1909–1910 well in advance of the foundation of quantum mechanics. The results set out below are essentially contained in these works.
\[ \int_0^b |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |F_b(\lambda)|^2 \, d\rho_b(\lambda). \]

It turns out that the family of functions \((\rho_b(\lambda))_{b>0}\) is uniformly bounded on each finite interval. This allows us to apply the theorem of Helly for \(b \to \infty\) (and arbitrary variation \(\beta\)) and to establish the following basic result.

**Theorem 1.** For the Schrödinger operator \(L[y] = -y'' + q(x)y\) with continuous real potential \(q(x)\) and boundary condition \(y'(0) - hy(0) = 0\) \((h \in \mathbb{R} \cup \{\infty\})\) there exists a non-decreasing spectral function \(\rho(\lambda) (-\infty < \lambda < \infty)\) such that for any \(f \in L^2(0, \infty)\) the generalised Fourier transform

\[ F(\lambda) = \lim_{b \to \infty} \int_0^b f(x)\psi(x, \lambda) \, dx \]

exists in \(L^2(-\infty, \infty)\), the Weyl inversion formula\(^{31}\)

\[ f(x) = \lim_{N \to \infty} \int_{-N}^N F(\lambda)\psi(x, \lambda) \, d\rho(\lambda) \] \hspace{1cm} \text{(58*)} \]

holds (convergence in \(L^2(0, \infty)\)) and Parseval's equality

\[ \int_0^{\infty} |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |F(\lambda)|^2 \, d\rho(\lambda) \]

is valid.

In the simplest case of \(q(x) = 0\) and boundary condition \(y'(0) = 0\) we have \(\psi(x, \lambda) = \cos \sqrt{\lambda} x\). Under the condition \(\psi'(b, \lambda) = 0\) we obtain:

\[ \lambda_n = \left( \frac{m \pi}{b} \right)^2, \quad \sigma_n^2 = \frac{b}{2} \quad (n = 0, 1, 2, \ldots); \]

\[ \rho_b(\lambda) = \frac{2}{b} \left( \left\lfloor \frac{h}{\sigma_b} \right\rfloor + 1 \right) \]

(\([\cdot]\) denotes the integral part) for \(\lambda \geq 0\) while \(\rho_b(\lambda) = 0\) for \(\lambda < 0\). Letting \(b \to \infty\) we obtain the unique limit function:

\(\rho(\lambda) = \frac{2}{\pi} \sqrt{\lambda} \quad (\lambda \geq 0); \quad \rho(\lambda) = 0 \quad (\lambda < 0).\)

After the substitution \(\sqrt{\lambda} = \mu\) the eigenfunction expansion takes the form

\[ f(x) = \lim_{N \to \infty} \frac{2}{\pi} \int_0^N \Phi(\mu) \cos \mu x \, d\mu, \]

where

\(^{31}\) Although the inversion formula is obtained heuristically from (57) on letting \(b \to \infty\), its rigorous derivation requires a deeper argument.
Chapter 1. Classical Concrete Problems

\[ \Phi(\mu) = \lim_{b \to \infty} \int_{0}^{b} f(x) \cos \mu x \, dx \]

(both defined in \( L^2(0, \infty) \)). Plancherel’s theorem has now been obtained for the case of an even function\(^{32}\).

In general the spectral function \( \rho(\lambda) \) is not unique. The situation here is analogous to the power moment problem.

We shall consider \( \psi(x, \zeta) \) for complex \( \zeta \) (\( \text{im} \, \zeta \neq 0 \)).

**Lemma.** If \( \psi(x, \zeta_0) \in L^2(0, \infty) \) for some \( \zeta_0 \) (\( \text{im} \, \zeta_0 \neq 0 \)) then \( \psi(x, \zeta) \in L^2(0, \infty) \) for all \( \zeta \) (\( \text{im} \, \zeta \neq 0 \)).

The quantity

\[ r(\zeta) = \frac{1}{2|\text{im} \, \zeta|} \left( \int_{0}^{\infty} |\psi(x, \zeta)|^2 \, dx \right)^{-1} \]

is called the Weyl radius at the point \( \zeta \). The lemma implies that on the set \( \text{im} \, \zeta \neq 0 \) we have the Weyl alternative: either \( r(\zeta) = 0 \) for all \( \zeta \) or \( r(\zeta) > 0 \) for all \( \zeta \). The Weyl disk or point is defined by the inequality

\[ \int_{0}^{\infty} |w\psi(x, \zeta) + \theta(x, \zeta)|^2 \, dx \leq -\frac{\text{im} \, w}{\text{im} \, \zeta}, \quad (59) \]

where \( \theta(x, \zeta) \) is the solution of the equation \( L[y] = \zeta y \) for the initial conditions \( \theta(0, \zeta) = -\sin \alpha \), \( \theta'(0, \zeta) = \cos \alpha \). The non-emptiness of the set of values of \( w \) defined by the inequality \( (59) \) is established by passage to the limit from the finite interval. Now the inequality

\[ \int_{0}^{b} |w\psi(x, \zeta) + \theta(x, \zeta)|^2 \, dx \leq -\frac{\text{im} \, w}{\text{im} \, \zeta}, \quad (60) \]

is satisfied if the solution \( \phi_w(x, \zeta) = w\psi(x, \zeta) + \theta(x, \zeta) \) is such that the Wronskian of the pair \( \phi_w, \overline{\phi_w} \) equals zero at the point \( b \). Such solutions clearly exist and therefore \( (60) \) defines a certain disk \( C_b \). Obviously \( C_{b_1} \subset C_{b_2} \) if \( b_1 > b_2 \). Thus \( C_{\infty} = \bigcup_{b} C_b \) is non-empty and this is also the disk or point of \( (59) \). This argument provides the reason for calling \( C_{\infty} \) the limit disk (or point). Its principal result is

**Theorem 2.** If \( \text{im} \, \zeta \neq 0 \) the equation \( L[y] = \zeta y \) has a solution belonging to \( L^2(0, \infty) \).

In the disk case, all solutions of the equation \( L[y] = \zeta y \) belong to \( L^2(0, \infty) \) while in the point case only multiples of \( \phi(x, \zeta) \) belong to \( L^2(0, \infty) \). Incidentally, it is clear from this that the realisation of one or other case of the Weyl alternative does not depend on the boundary condition at zero, i.e. it depends only on the potential. For example, if the potential is bounded from below or if it only admits

\(^{32}\)It is appropriate to note that even Fourier deduced his integral formulae heuristically by passage to the limit from series.


a bound \( q(x) > -cx^2 \) (\( c \) a positive constant) then the point case occurs (Hartman-Wintner, 1949).

**Theorem 3.** The spectral function is unique in the point case but not unique in the disk case.

The following fundamental definition is only meaningful in the point case.

The set of points of increase of the spectral function \( \rho(\lambda) \) is called the *spectrum* of the Schrödinger operator \( L[y] = -y'' + g(x)y \) with boundary condition \( y'(0) - hy(0) = 0 \). If \( \mu \) is a point in the spectrum, the solution \( \psi(x, \mu) \) is called a *generalised eigenfunction* corresponding to the point \( \mu \).

In this sense (58*) gives an expansion with respect to the generalised eigenfunctions of the Schrödinger operator. We have a proper *eigenfunction* \( \psi(x, \mu) \) (i.e. \( \psi(\cdot, \mu) \in L^2(0, \infty) \)) if and only if \( \mu \) is a jump point of the spectral function. The set of eigenvalues is called the *discrete spectrum* of the operator. This term is connected with the fact that any isolated point of the spectrum is an eigenvalue. The set of non-isolated points of the spectrum is called the *continuous spectrum*.

**Theorem 4.** If \( q(x) \to +\infty \) (\( x \to \infty \)) the spectrum of the Schrödinger operator is purely discrete.

This test of H. Weyl was substantially improved by A.M. Molchanov (1953).

**Theorem 5.** A necessary and sufficient condition for the discreteness of the spectrum of the Schrödinger operator with potential \( q(x) \) which is bounded below is that

\[
\int_x^{x+a} q(\xi) \, d\xi \to +\infty \quad (x \to \infty)
\]

for each \( a > 0 \).

In contrast to this we have

**Theorem 6.** If \( \int_0^\infty |q(x)| \, dx < \infty \) then the positive semiaxis \( \lambda \geq 0 \) is contained in the continuous spectrum and does not contain the discrete spectrum while in the negative semiaxis the spectrum is purely discrete.

### 2.13. Almost-Periodic Functions.

The class \( AP \) of almost-periodic functions (a.p.f.s) was introduced and studied in 1925 by Bohr.

---

33 As a result of (58) and (58*) the spectral function provides a measure on the spectrum with respect to which the system of generalised eigenfunctions is orthonormal in the sense that

\[
\int_{-\infty}^{\infty} \psi(x, \lambda)\psi(y, \lambda) \, d\rho(\lambda) = \delta(x - y).
\]

34 Among these may be eigenvalues.

35 In a more detailed notation, \( AP(R) \).

36 Following initial investigations of Bohl and Esclangon.
A complex continuous function \( f(t) (t \in \mathbb{R}) \) is said to be an a.p.f. if for any \( \varepsilon > 0 \) there exists a relatively dense set of \( \varepsilon \)-almost periods, i.e. numbers \( \tau \) such that \( \sup_t |f(t + \tau) - f(t)| < \varepsilon \). A set \( M \subset \mathbb{R} \) is said to be relatively dense if there exists \( l > 0 \) such that each interval of length \( l \) contains a point of \( M \). Bochner (1927) showed that the definition of an a.p.f. is equivalent to the following: from every sequence of translations \( (f(t + \tau_n)) \) it is possible to extract a subsequence which converges uniformly on the whole axis. This definition opens up the possibility of a more transparent construction of the theory of a.p.fs.

All periodic functions, in particular all exponentials \( e^{i\lambda t} (\lambda \in \mathbb{R}) \), are a.p.fs. The class \( AP \) is a linear space. Therefore all exponential polynomials \( \sum \chi \lambda_i e^{i\lambda_i t} \) are a.p.fs. The uniform limit on the whole axis of a sequence of a.p.fs is also an a.p.f. Thus any function which allows an arbitrarily close uniform approximation by exponential polynomials is an a.p.f.

**Theorem of Bohr.** Any a.p.f. is the uniform limit on the whole axis of a sequence of exponential polynomials.

This analogue of Weierstrass's theorem is closely connected with the theory of Fourier series of a.p.fs where an analogue of \( L^2 \)-theory plays a key role.

**Lemma.** For any a.p.f. \( f \) the mean

\[
m(f) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t) \, dt
\]

exists. If \( f \geq 0 \) and \( m(f) = 0 \) then \( f = 0 \).

This allows us to introduce on the space \( AP \) the scalar product \( (f, g) = m(f \overline{g}) \) (\( AP \) is not only a linear space but also an algebra).

Let us consider the continuous family of exponentials: \( e_{\lambda}(t) = e^{i\lambda t} (\lambda \in \mathbb{R}) \). It is orthonormal: \( (e_{\lambda}, e_{\mu}) = \delta_{\lambda\mu} \). The following definition therefore arises naturally: \( c_{\lambda}(f) = (f, e_{\lambda}) \) is called a Fourier coefficient of the a.p.f. \( f \). Following on from this we introduce formally the Fourier series

\[
f(t) \sim \sum_{\lambda} c_{\lambda}(f) e^{i\lambda t}.
\]

In fact, this series contains no more than countably many non-zero terms, since Bessel's inequality

\[
\sum_{k=1}^{n} |c_{\lambda_k}(f)|^2 \leq (f, f)
\]

holds for any collection \( \{\lambda_1, \ldots, \lambda_n\} \). The set of those \( \lambda \) for which \( c_{\lambda}(f) \neq 0 \) is called the Bohr spectrum of the function \( f \).

---

37 In the terminology of physics this is the superposition of harmonic oscillations with arbitrary frequencies. In particular, the coordinate functions describing small oscillations of conservative mechanical systems are of this type.
§ 3. Theory of Approximation

The theorem of Bohr reduces to Parseval's equality

$$\sum_{k} |c_k(f)|^2 = (f, f),$$

which means that the Fourier series of an a.p.f. $f$ converges to $f$ in the sense of Bohr mean-square: if we put $\|f\| = \sqrt{(f, f)}$ and introduce the distance $d(f_1, f_2) = \|f_1 - f_2\|$ then $d(f, S_n) \to 0$ as $n \to \infty$, where $S_n$ is the $n$th partial sum of the Fourier series of the function $f$ for an arbitrarily chosen enumeration of the non-zero terms. It is also possible to sum the Fourier series of an a.p.f. $f$ to $f$ using a method of Fejér type (Bochner, 1927).

§ 3. Theory of Approximation

3.1. Chebyshev Approximations. According to the theorem of Weierstrass any function $f$ which is continuous on the interval $[a, b]$ allows an arbitrarily close uniform approximation by means of polynomials. The problem therefore arises naturally of determining amongst all polynomials of degree no higher than $n$ that polynomial $P$ for which the deviation

$$d(f, P) = \max_{a \leq x \leq b} |f(x) - P(x)|$$

is as small as possible. The formulation of this problem (for a real function) is due to P.L. Chebyshev (1853) who started from considerations connected with the theory of mechanisms. Later on this approach developed into an independent and deep area of mathematical analysis.

The existence of a polynomial of best approximation is obvious, since the function $d(f, P)$ is continuous on the space of polynomials of degree $\leq n$ and tends to $+\infty$ at infinity. The problem of uniqueness is rather more subtle and is closely linked with the problem of effectively constructing the polynomial of best approximation.

Theorem of Chebyshev. In order that a polynomial $P(x)$ (deg $P \leq n$) be the polynomial of best approximation for a real-valued continuous function $f(x)$ ($a \leq x \leq b$), it is necessary and sufficient that there exist a collection of points $x_1 < x_2 < \cdots < x_{n+2}$ in $[a, b]$ (Chebyshev alternant) at which the remainder $R(x) = f(x) - P(x)$ satisfies the conditions:

1) $|R(x_k)| = d(f, P) \ (1 \leq k \leq n + 2);$
2) $\text{sgn} \ R(x_{k+1}) = -\text{sgn} \ R(x_k) \ (1 \leq k \leq n + 1).$

1 On the space $C(X)$ of continuous functions on any compact set $X$ the natural metric (corresponding to uniform convergence and therefore called uniform) is defined as $d(f_1, f_2) = \|f_1 - f_2\|$, where $\|f\| = \max_{x \in X} |f(x)|$. Thus the deviation $d(f, P)$ is the distance from $P$ to $f$.

2 i.e. long before the appearance of Weierstrass' theorem.

3 The proof of this theorem was developed gradually and a complete version did not appear until the work of Borel (1905).
Corollary. The polynomial of best approximation for given $f$, $n$ is unique.

Proof. The semi-sum $S$ of two polynomials of best approximation $P$, $Q$ is also a polynomial of best approximation. It is easy to show that the Chebyshev alternant for $S$ acts in the same way for $P$ and $Q$. It follows from this that $P$ and $Q$ coincide at the $n + 2$ points of the alternant. □

A similar result holds for best uniform approximations of continuous periodic functions by means of trigonometric polynomials, the only difference being that for a given degree $n$ the Chebyshev alternant consists of $2n + 2$ points in a period. We can deduce this from the previous result by substituting $x = \tan \frac{\theta}{2}$.

Let us consider some examples.

Example 1. P.L. Chebyshev posed and solved the following problem: among all polynomials of degree $n$ with leading coefficient $a_0 = 1$ find the one whose maximum modulus on $[-1, 1]$ is least (the polynomial deviating least from zero). It was shown by him to be the polynomial

$$T_0(x) = 1 \quad (n = 0), \quad T_n(x) = \frac{1}{2^{n-1}} \cos(n \arccos x) \quad (n \geq 1);$$

these were subsequently called the Chebyshev polynomials. The verification of this answer using Chebyshev's theorem causes no problem if we note that the desired polynomial has the form $x^n - Q_{n-1}(x)$ where $Q_{n-1}(x)$ is the polynomial of best approximation (deg $Q_{n-1} \leq n - 1$) for $x^n$ on $[-1, 1]$.

It is interesting to note that the Chebyshev polynomials are orthogonal on $(-1, 1)$ for the weight $1/(1 - x^2)$.

Example 2. We consider Weierstrass function

$$W(\theta) = \sum_{k=0}^{\infty} a^k \cos(b^k \theta)$$

where $0 < a < 1$ and $b$ is an odd integer $> 1$. This is a continuous $2\pi$-periodic function which does not have a derivative at any point if $ab > 1$. It is given by its Fourier series which converges uniformly to it. It can be shown that the partial sums of this series are the best uniform approximations for $W$ by means of trigonometric polynomials. This situation is encountered extremely rarely, in contrast to best approximations in mean square where it is always the case.

For any complex $2\pi$-periodic continuous function $f$ we shall denote by $E_n(f)$ the deviation from $f$ of trigonometric polynomials$^4$ of degree $\leq n$. From the formula for the partial sum of a Fourier series we obtain the bound

$$\max_{0 \leq x \leq 2\pi} |f(x) - S_n(x)| \leq (L_n + 1)E_n(f), \quad (61)$$

$^4$The same notation is also used in an algebraic situation and also in subsequent generalisations which cover both situations.
where

\[ L_n = \int_{-\pi}^{\pi} |D_n(x)| \, dx \]

is the Lebesgue constant (the \( L^1 \)-norm of the Dirichlet series). Weierstrass' theorem shows that \( E_n(f) \to 0 \) as \( n \to \infty \). At the same time \( L_n \to \infty \) but grows slowly. In fact as Lebesgue showed (1910),

\[
\frac{4}{\pi^2} \ln n \leq L_n \leq \frac{4}{\pi^2} \ln n + O(1).
\]

Inequality (61) can therefore be useful as a bound for the rate of convergence of a Fourier series. This is the case for example with smooth functions since we have

**Theorem of Jackson.** Let \( f \) be a function of class \( C^r \) on the circle and let \( \omega_r(\delta) \) be the modulus of continuity of its \( r \)th derivative \( (r \geq 0) \) \( (\omega_0(\delta) \equiv \omega(\delta), \) the modulus of continuity of \( f) \). Then

\[
E_{n-1}(f) \leq B_r \omega_r \left( \frac{1}{n} \right) n^{-r},
\]

(63)

where the constant \( B_r \) depends only on \( r \). In particular, if \( f^{(r)} \in \text{Lip} \alpha \) then

\[
E_n(f) = O(n^{-r-\alpha}).
\]

(64)

A bound of type (63) was also obtained by Jackson in the algebraic situation. In the class of functions which are analytic on the interval \([a, b]\) we have the bound of S.N. Bernshtein

\[
\lim_{n \to \infty} \sqrt[n]{E_n(f)} < 1,
\]

(65)

and moreover if \( f \) is an entire function

\[
\lim_{n \to \infty} \sqrt[n]{E_n(f)} = 0.
\]

(66)

To S.N. Bernshtein is due a series of deep results, the main point of which lies in the fact that the rate of convergence of the deviation \( E_n(f) \) to zero determines the class of smoothness of the function \( f \). We formulate in particular those theorems which provide converses to the bounds mentioned above.

**Theorem 1.** If (64) is satisfied for a function \( f \in C[a, b] \) then \( f \in C^r[a, b] \) and \( f^{(r)} \) belongs to the Lipschitz class with exponent \( \alpha \) on each interval \([a_1, b_1]\) \( \subset (a, b) \).

**Theorem 2.** If (65) is satisfied for a function \( f \in C[a, b] \) then \( f \) is analytic on \([a, b] \) and in the case of (66) it is an entire function.

Similar theorems hold in the trigonometric situation.

One of the proofs of Theorem 1 is based on the following remarkable inequality of Bernshtein: if \( T(\theta) \) is a trigonometric polynomial of degree \( n \) then

\[
\max_{-\pi \leq \theta \leq \pi} |T'(\theta)| \leq n \max_{-\pi \leq \theta \leq \pi} |T(\theta)|;
\]
further if \( f(z) \) is an entire function of exponential type \( \sigma \) which is bounded on the real axis, then
\[
\sup_{-\infty < x < \infty} |f'(x)| \leq \sigma \sup_{-\infty < x < \infty} |f(x)|
\]
(equality holds for the functions \( f(z) = e^{\pm i\sigma z} \)). This result can be generalised considerably (B.Ya. Levin, 1951).

The inverse problem of approximation, which was also formulated and resolved by S.N. Bernshtein (1938), consists of constructing a continuous function \( f \) with a given sequence \( (E_n)_{n=0}^{\infty} \) of deviations. The obvious necessary condition that \( (E_n)_{n=0}^{\infty} \) tend to zero monotonically turns out also to be sufficient.

3.2. Chebyshev and Markov Systems. A system \( (u_k)_{k=0}^{\infty} \) of real or complex continuous functions on a topological space \( X \) is called a Chebyshev system if each linear combination of these functions, which is not identically zero, has no more than \( n \) roots. This class of systems and the related class of Markov systems (see below) arise very naturally in many questions in the theory of approximation and its applications. The standard examples of Chebyshev systems are:

\[
1, z, \ldots, z^n \quad (z \in \mathbb{C});
\]

\[
1, \cos x, \sin x, \ldots, \cos nx, \sin nx \quad (0 \leq x < 2\pi).
\]

A further example is the system of exponentials
\[
e^{i\lambda_0 x}, \ldots, e^{i\lambda_n x} \quad (-\infty < x < \infty)
\]
with real, pairwise distinct exponents \( \lambda_0, \ldots, \lambda_n \). Using a logarithmic substitution we can transform it into the Chebyshev system
\[
x^{\lambda_0}, \ldots, x^{\lambda_n} \quad (x > 0).
\]

The inequality
\[
\Delta \left( \begin{array}{c} u_0 \ldots u_n \\ x_0 \ldots x_n \end{array} \right) = \det(u_k(x_i))_{i,k=0}^{n} \neq 0
\]
for all pairwise distinct \( x_0, \ldots, x_n \) provides a criterion under which \( (u_k)_{k=0}^{\infty} \) is a Chebyshev system. Clearly any Chebyshev system is linearly independent.

Let \( X \) be any compact space. Given a system \( (u_k)_{k=0}^{\infty} \subset C(X) \) and \( f \in C(X) \) there clearly exists a linear combination \( \sum \alpha_k u_k \) with least deviation from \( f \) in the uniform metric.

**Tonelli-Haar-Kolmogorov Theorem.** Let \( (u_k)_{k=0}^{\infty} \subset C(X) \) be a linearly independent system. For any \( f \in C(X) \) there is a unique linear combination \( P = \sum \lambda_k u_k \) with least deviation from \( f \) in the uniform metric if and only if \( (u_k)_{k=0}^{\infty} \) is a Chebyshev system.

Interestingly, it is far from true that Chebyshev systems (with \( n > 0 \)) exist on each compact space (Mairhuber, 1956; Sieklucki, 1958).

**Theorem.** If a real Chebyshev system (with \( n > 0 \)) exists on the compact space \( X \) then \( X \) can be embedded homeomorphically in the circle.
A similar result holds in the complex situation (Schoenberg-Yang, 1961).

**Theorem.** If $X$ is a finite polyhedron on which there exists a complex Chebyshev system (with $n > 0$) then $X$ can be embedded homeomorphically in the plane.

Suppose that $X$ is a compact space, $(u_k)_0^n \subset C(X)$ is a complex Chebyshev system and $f \in C(X)$. The following criterion for best approximation is due to A.N. Kolmogorov (1948).

**Theorem.** The linear combination $P = \sum_{k=0}^n \lambda_k u_k$ deviates least from $f$ in the uniform metric, if and only if the inequality

$$\min_{x \in M_p} \left| f(x) - P(x) \right| = 0$$

holds for all linear combinations $F = \sum_{k=0}^n \alpha_k u_k$, where

$$M_p = \{x: |f(x) - P(x)| = \max\}.$$

**Corollary.** If the compact space $X$ contains at least $n + 2$ points then the set $M_p$ for the best approximation $P$ also contains at least $n + 2$ points.

In the case $X = [a, b]$ the best approximation by linear combinations of a real Chebyshev system can be characterised as before by the existence of a Chebyshev alternant (S.N. Bernshtein, 1937).

The following theorem of Jackson shows that the Chebyshev property of a system is connected with the properties of best approximations not only in the uniform metric but also in the $L^1$-metric:

if $(u_k)_0^n$ is a real Chebyshev system on $[a, b]$ then for any real function $f \in C[a, b]$ the linear combination $P = \sum \lambda_k u_k$ which deviates least from $f$ in the $L^1$-metric is unique.

In the class $L^1(a, b)$ we do not have uniqueness of the best approximation no matter what the system $(u_k)_0^n \subset L^1(a, b)$.

A certain strengthening of the requirements made on the system leads to the possibility of effectively calculating the least deviation in the $L^1$-metric.

A system $(u_k)_0^n (v \leq \infty)$ of real or complex continuous functions on a topological space $X$ is called a Markov system if each subsystem $(u_k)_0^n (0 \leq n < v + 1)$ is a Chebyshev system. Let $(u_k)_0^n$ be a real Markov system on the interval $[a, b]$. We form the combinations

$$v_n = u_0, \quad v_n = u_n - \sum_{k=0}^{n-1} \alpha_k u_k$$

where for $n \geq 1$ these have least deviation from zero in the $L^1$-metric. It can be shown that the function $v_n$ has exactly $n$ roots $x_{n1} < \cdots < x_{nn}$ in $(a, b)$.

Let $f \in C[a, b]$. We put $P_n = \sum_{k=0}^n \lambda_k u_k$, where the $\lambda_k$ are defined by the system of equations

$$P_n(x_{n+1,i}) = f(x_{n+1,i}) \quad (1 \leq i \leq n + 1).$$

**Theorem of Markov.** If the remainder $f(x) - P_n(x)$ alternates in sign as $x$ takes on the successive values $x_{n+1,1}, \ldots, x_{n+1,n+1}$, then for the $L^1$-metric $P_n$ has least
deviation from \( f \) among all combinations of the form \( \sum_{k=0}^{n} \alpha_k u_k \). Further the least deviation is equal to
\[
\left| \int_{a}^{b} f(x) \, \text{sgn} \, v_{n+1}(x) \, dx \right|.
\]

A.N. Korkin and E.I. Zolotarev (1873) constructed a polynomial of degree \( n \) with leading coefficient 1 having least deviation from zero in \( L^1(-1, 1) \). It proved to be the Chebyshev polynomial of the second kind:
\[
\tilde{T}_n(x) = \frac{1}{2^n} \frac{\sin(n+1) \arccos x}{\sqrt{1-x^2}}.
\]
The corresponding least deviation is equal to \( \frac{1}{2^{n-1}} \). These results can be deduced from Markov's theorem.

We note that the Chebyshev polynomials of the second kind are orthogonal on \((-1, 1)\) with weight \( \sqrt{1-x^2} \).

3.3. The Chebyshev-Markov Problem. For any system of real continuous functions \( (u_k)_{k=0}^{\nu} \) on the interval \([a, b] \) and any system of real numbers \( (c_k)_{k=0}^{\nu} \) we can consider the moment problem which consists of finding a non-decreasing function \( \sigma(x) \) \((a \leq x \leq b)\) satisfying the conditions
\[
\int_{a}^{b} u_k(x) \, d\sigma(x) = c_k \quad (0 \leq k < \nu + 1).
\]

An obvious necessary condition for the existence of a solution is that for any system of real numbers \( (x_k)_{k=0}^{n} \) \((n < \nu + 1)\) such that
\[
\sum_{k=0}^{n} \alpha_k u_k(x) \geq 0 \quad (a \leq x \leq b), \quad (67)
\]
the inequality
\[
\sum_{k=0}^{n} \alpha_k c_k \geq 0 \quad (68)
\]
should hold. It can be shown that if \( u_0(x) > 0 \) \((a \leq x \leq b)\) then this condition is sufficient for the solvability of the moment problem under consideration. Suppose that it is satisfied and let \( \nu = n < \infty \). The Chebyshev-Markov problem consists of finding for a given continuous function \( \Omega(x) \geq 0 \) \((a \leq x \leq b)\) and a given point \( \xi \in (a, b) \) the greatest value of the integral
\[
\int_{a}^{\xi+0} \Omega(x) \, d\sigma(x)
\]
and the least value of the integral
\[ \int_a^{\xi-0} \Omega(x) \, d\sigma(x) \]
on the set of all solutions of the moment problem just described\(^5\).

Let us assume that \((u^k_0)\) is a normalised Markov system in the sense that \[ \Delta(u_0 \ldots u_m) > 0, \]
if \(a = x_0 < \cdots < x_m < b (0 \leq m \leq n)\). We assume further that the function \(\Omega\) associated with this system does not disturb its 'Markovicity' even if in the following sense:
\[ \Delta(u_0 \ldots u_{m-1} \Omega) \geq 0 \quad (0 \leq m \leq n + 1). \]

All these conditions are satisfied, in particular, if \(u_k(x) = x^k (0 \leq k \leq n), \Omega \in C^{n+1}(a, b]\) and \(\Omega^{n+1}(x) \geq 0 (a < x < b)\). With regard to the given moments \(c_k (0 \leq k \leq n)\) we assume that the inequality (68) is strict for all non-zero collections \((x_k)\) which satisfy condition (67). Then we can show that among the solutions to the moment problem there exists a unique piecewise-constant function \(\sigma_\xi\) which has a jump at the point \(\xi\) and satisfies the condition \(2p + q \leq n + 2\) where \(p\) and \(q\) are respectively the number of jump points in the open interval \((a, b)\) and on its boundary.

The Chebyshev-Markov inequalities assert that
\[ \int_a^{\xi+0} \Omega(x) \, d\sigma(x) \leq \int_a^{\xi+0} \Omega(x) \, d\sigma_\xi(x) \]
and
\[ \int_a^{\xi-0} \Omega(x) \, d\sigma(x) \geq \int_a^{\xi-0} \Omega(x) \, d\sigma_\xi(x). \]

Thus the function \(\sigma_\xi(x)\) is the desired solution of both extremal problems and moreover this solution does not depend on \(\Omega\).

The piecewise-constant solutions of the moment problem naturally lead to quadrature formulae of Gaussian type. In fact if \(x_1 < \cdots < x_r\) form the set of jump points of this solution with \(p_1, \ldots, p_r\) the corresponding jumps, then the approximate formula
\[ \int_a^b f(x) \, d\sigma(x) \approx \sum_{j=1}^n \rho_j f(x_j) \]
is exact for all linear combinations \(f = \sum_{k=0}^n \lambda_k u_k\) and solutions \(\sigma\) of the given moment problem. The function \(\sigma\) can be assigned a priori, then the moment sequence is calculated and the nodes \(x_j\) and weights \(\rho_j (1 \leq j \leq r)\) are already

\(^5\) The first work of P.L. Chebyshev (1874) was devoted to this case with \(\Omega(x) \equiv 1\).
defined by it. In particular in the Gaussian case, i.e. \( u_k(x) = x^k \) \((0 \leq k \leq 2r - 1; a \leq x \leq b)\), we have the roots \( x_1, \ldots, x_r \) of the orthogonal polynomial \( P_r(x) \) as nodes, while the weights (the Christoffel coefficients) are calculated from the formula:

\[
\rho_j = \frac{Q_r(x_j)}{P'_r(x_j)} = \text{Res}_{x=x_j} \frac{Q_r(x)}{P_r(x)} \quad (1 \leq j \leq r),
\]

where \( Q_r(x) \) is the associated polynomial. The corresponding piecewise-constant function is the solution of the Chebyshev-Markov problem if \( \xi = x_l \) for some \( l \).

If in addition we put \( \Omega(x) \equiv 1 \), then the Chebyshev-Markov inequalities take the form

\[
\int_{a}^{x_l} d\sigma(x) \leq \sum_{j=1}^{l} \rho_j, \quad \int_{a}^{x_l} d\sigma(x) \geq \sum_{j=1}^{l-1} \rho_j.
\]

### 3.4. The \( L \)-problem of Moments.

Another interesting type of moment problem goes back to the investigations of A.A. Markov: it is required to find a real measurable function \( p(x) \) which satisfies the constraint

\[
|p(x)| \leq L \quad (a \leq x \leq b)
\]

(\( L \) is a given positive number) and has prescribed moments

\[
\int_{a}^{b} u_k(x)p(x) \, dx = c_k \quad (0 \leq k \leq n)
\]

with respect to the given real system \( (u_k)_n \in L^1(a, b) \). We call this statement of the question the \( L \)-problem of moments. The criterion for its solvability is closely connected with least deviations from zero. It is immediately clear that if a solution to the problem (69)–(70) exists, then for any collection of numbers \( (\alpha_k)_n \) satisfying the condition

\[
\sum_{k=0}^{n} \alpha_k c_k = 1,
\]

we have the inequality

\[
L \left| \int_{a}^{b} \left| \sum_{k=0}^{n} \alpha_k u_k(x) \right| \, dx \right| \geq 1.
\]

Thus if

\[
l = \left( \min_{(\alpha_k)} \int_{a}^{b} \left| \sum_{k=0}^{n} \alpha_k u_k(x) \right| \, dx \right)^{-1}
\]

where \( (\alpha_k) \) satisfies (71), we obtain \( L \geq l \) as a necessary condition for the solvability of the problem (69)–(70).

---

It is appropriate to note here that in any \( L^2(I) \) the roots of the orthogonal polynomials are simple and lie inside the interval \( I \).
N.I. Akhiezer and M.G. Krejn (1938) obtained the following result.

**Theorem.** Suppose the system \((u_k)_0\) has the property that for each non-zero linear combination \(\sum \alpha_k u_k\) its set of zeros has measure zero. The \(L\)-problem (69)–(70) has a solution if \(L \geq l\). This solution is essentially\(^7\) unique if and only if \(L = l\), in which case it has the form

\[ p(x) = L \sgn \hat{Q}(x), \]

where \(\hat{Q}\) is a solution of the dual extremal problem\(^8\)

\[ \int_a^b |Q(x)| \, dx = \min \]

on the set of functions of the form

\[ Q = \sum_{k=0}^n \alpha_k u_k \left( \sum_{k=0}^n \alpha_k c_k = 1 \right). \]

The solution for the power and trigonometric cases of the \(L\)-problem of moments can be described effectively in a certain sense.

### 3.5. Interpolation and Quadrature Processes.

The exact solution of the best approximation problem lends itself relatively rarely to an effective description while an approximate solution can require complicated calculations.

The more traditional method of approximating functions, namely interpolation, is effective and simple from the computational point of view, although the approximations provided by it can turn out to be very coarse.

**Interpolation** of a function consists of constructing those linear combinations of the given system \((u_k)_0\) (e.g. polynomials) which take the same values as \(f\) at the given points (interpolating points). A more refined procedure also requires equality of derivatives up to a given order (with corresponding assumptions about their existence) so that, for example, from this point of view Taylor's formula is an interpolation formula. We restrict ourselves below to interpolation with simple interpolating points, i.e. without regard to derivatives.

If \((u_k)_0^v (v < \infty)\) is a Markov system on the space \(X\) then for any interpolating points \(x_j \in X\) and for any numerical values \(y_j (0 \leq j \leq n < v + 1)\) there is a unique linear combination \(P = \sum_{k=0}^n \lambda_k u_k\) such that \(P(x_j) = y_j (0 \leq j \leq n)\). It can be expressed in the following form:

\[ P(x) = \sum_{k=0}^n y_k \Phi_k(x), \quad (72) \]

where \(\Phi_k\) is the linear combination of the functions \(u_0, \ldots, u_n\) which satisfies the conditions \(\Phi_k(x_j) = \delta_{kj}\) (in this sense the systems of functions \((\Phi_k)_0^v\) and interpolating points \((x_j)_0^v\) are biorthogonal).

\(^7\)i.e. up to values on a set of measure zero.

\(^8\)The solution \(\hat{Q}\) is not unique but the signs of any two solutions coincide almost everywhere.
In the case \( u_k(x) = x^k \) \((0 \leq k \leq n)\) on the interval \([a, b]\) we have

\[
\Phi_k(x) = \frac{\omega(x)}{\omega'(x_k)(x - x_k)},
\]

where

\[
\omega(x) = \prod_{j=0}^{n} (x - x_j),
\]

and (72) becomes the Lagrange interpolation formula:\n
\[
P(x) = \sum_{k=0}^{n} y_k \frac{\omega(x)}{\omega'(x_k)(x - x_k)}.
\]

If this polynomial is expressed in terms of the basis

\[
1, (x - x_0), (x - x_0)(x - x_1), \ldots, (x - x_0)(x - x_1)\ldots(x - x_{n-1})
\]

of the space of polynomials of degree \( \leq n \), we obtain Newton's interpolation formula, the advantage of which is that the introduction of a new interpolating point requires only the addition of a new term to the formula without having to recalculate the previous ones. Therefore, if we are given an infinite sequence of interpolating points \((x_j)_{n}^{\infty}\) and corresponding values \((y_j)_{n}^{\infty}\), Newton's interpolation series \((N.i.s.)\)

\[
\sum_{k=0}^{\infty} \frac{a_k}{k!} \prod_{j=0}^{k-1} (x - x_j)
\]

is defined. If N.i.s. converges even only pointwise on some set containing all the interpolating points, then its sum \( f \) is a solution of the interpolation problem

\[
f(x_j) = y_j \quad (j = 0, 1, 2, \ldots).
\]

The convergence of N.i.s. with some or other interpolating points in the complex plane (in particular, with the interpolating points \( x_j = j^{10} \)) has been studied in considerable detail. It is interesting that this tool can be used for the proof of the famous theorem of Lindemann that the numbers \( x \neq 0 \) and \( e^x \) cannot be simultaneously algebraic (for \( x = 1 \) this yields the transcendental nature of the number \( e \) and for \( x = \pi i \) that of the number \( \pi \)) (A.O. Gel'fond, 1929).

Let \((u_k)_{n}^{\infty}\) be a Markov system and for each \( n = 0, 1, 2, \ldots \) let interpolating points \( x_{nj} \) \((0 \leq j \leq n)\) be given. The sequence of linear combinations

\[\text{A similar (but of course more complicated) formula resolving the problem of interpolation with multiple interpolation points was constructed by Hermite.}\]

\[\text{In this case N.i.s. converges or diverges simultaneously with the Dirichlet series}\]

\[
\sum_{k=0}^{\infty} \frac{(-1)^k a_k}{k^x}
\]

and the same holds for absolute convergence.
$P_n = \sum_{k=0}^{n} \alpha_{nk} u_k$

defined by the conditions $P_n(x_{nj}) = f(x_{nj})$ ($0 \leq j \leq n$) is called an interpolation process for the given function $f$. In the case where $x_{nj}$ does not depend on $n$ ($n \geq 1$) we speak of an interpolation process of Newton type.\(^{11}\)

One of the fundamental questions relating to interpolation processes is that of convergence (in some sense or other) of the sequence $(P_n)\_n$ to the function $f$. A positive resolution of this question depends on the distribution of the interpolating points and the properties of the function $f$. For example, if the Markov system $(\mathbf{u}_k)\_n$ is complete in $L^1(a, b)$ and consists of functions which are continuous on $[a, b]$ and if the $x_{nj}$ ($1 \leq j \leq n$) are the roots of the combinations

$$v_n = u_n - \sum_{k=0}^{n-1} \alpha_{nk} u_k \quad (n = 1, 2, \ldots)$$

which have least deviation from zero in the $L^1$-metric, then, under the conditions of Markov's theorem, the interpolation process for any $f \in C[a, b]$ converges to $f$ in the $L^1$-metric. In particular, the sequence of interpolating polynomials constructed on the roots of the Chebyshev polynomials of the second kind, i.e. on the interpolating points

$$x_{nj} = \cos \frac{j\pi}{n + 1} \quad (1 \leq j \leq n),$$

converges.

For uniform convergence of an interpolation process it is necessary to impose stringent restrictions on the function $f$. For example, if $f$ is an entire function the sequence of interpolating polynomials converges uniformly to $f$ on any interval $[a, b]$ for any distribution of the interpolating points. This follows easily from the formula for the remainder term

$$f(x) - P_n(x) = f^{(n+1)}(\xi_x) \frac{\omega(x)}{(n + 1)!} \quad (a < \xi_x < b),$$

which holds for all $f \in C^{n+1}[a, b]$. Moreover, as Faber showed (1914), for any distribution of interpolating points in $[a, b]$ there is a continuous function for which the sequence of interpolating polynomials is not uniformly bounded. For the function $f(x) = |x|$ on $[-1, 1]$ the sequence of interpolating polynomials corresponding to the equally spaced interpolating points

$$x_{nj} = \frac{2j - n}{n} \quad (0 \leq j \leq n)$$

is unbounded at each point other than $0, \pm 1$ (S.N. Bernshtein, 1916).

The quadrature process for the system of nodes $(x_{nj})_{0 \leq j \leq n}$ on the compact space $X$ with measure $\mu$ ($\mu(X) < \infty$) is defined by the sequence of approximate

\(^{11}\)In contrast to this we speak of an interpolation process of Lagrange type in the general case.
for the integral which is regarded as a functional of $f \in C(X)$. The coefficients $\rho_{nj}$ are found, for example, by requiring that the formula is exact for $f = u_k$ ($0 \leq k \leq n$), where $(u_k)_{0}^{\infty} \subset C(X)$ is a Markov system. Under this condition formulae (73) can be written in the form

$$\int_X f d\mu \approx \int_X P_n d\mu \quad (n = 0, 1, 2, \ldots),$$

where $(P_n)_{0}^{\infty}$ is the interpolation process for $f$ corresponding to the given system of nodes. Therefore if $(P_n)_{0}^{\infty}$ converges to $f$ in the $L^1$-metric, the quadrature process for $f$ also converges:

$$\lim_{n \to \infty} \sum_{j=0}^{n} \rho_{nj} f(x_{nj}) = \int_X f d\mu.$$
The interpolation process for \( f \) therefore converges if
\[
\lim_{n \to \infty} M_n E_n = 0.
\] (74)

The analogous condition for convergence of the quadrature process has the form
\[
\lim_{n \to \infty} \mu_n E_n = 0 \quad \left( \mu_n = \sum_{j=0}^{n} |\rho_{nj}| \right).
\]

The sufficient condition for convergence of a quadrature process which was formulated above follows immediately from this.

3.6. Approximation in the Complex Plane. If the continuous function \( f \) on the compact set \( X \subset \mathbb{C} \) allows arbitrarily close uniform approximation by means of polynomials, then \( f \) is analytic on the interior \( \text{int} \ X \) of the set \( X \) and it can be extended analytically onto all the bounded connected components of the complement \( \mathbb{C} \setminus X \). For a long time the sufficiency of these conditions remained an open problem. After early articles of Appel and Runge, the works of Walsh (1926), M.A. Lavrent'ev (1934) and M.V. Keldysh (1945) were fundamental landmarks in the path towards its resolution. The final positive solution of the problem of polynomial approximation is due to S.N. Mergelyan (1951) who proved the following theorem.

**Theorem.** If the complement of the compact set \( X \) is connected then each function \( f \) which is continuous on \( X \) and analytic on \( \text{int} \ X \) is the uniform limit of a sequence of polynomials.

In particular, if \( X \) is nowhere dense and its complement is connected then every continuous function on \( X \) can be approximated by polynomials with arbitrary degree of accuracy (theorem of Lavrent'ev). The simplest example of this situation is a Jordan arc (Walsh's case), in particular a closed bounded interval (Weierstrass's case). For an arbitrary compact set \( X \subset \mathbb{C} \) rational functions with poles outside of \( X \) are the natural means of approximating to continuous functions which are analytic on the interior of \( X \). The following criterion obtained by A.G. Vitushkin (1967) makes use of the idea of continuous analytic capacity of a set \( M \subset \mathbb{C} \):

\[
\alpha(M) = \sup_{f} \left| \left. \frac{\text{Res } f(z)}{z} \right| \right|_{z=\infty},
\] (75)

where \( f \) runs through the class of all functions which are continuous on the entire plane, analytic outside of some compact set \( Q_f \subset M \), bounded in modulus by unity and tend to zero at infinity.\(^\text{12}\)

\(^\text{12}\) The rather simpler notion of analytic capacity (denoted by \( \gamma(M) \)) is defined by the same formula (75) but on a larger class of functions \( f \), namely those which are defined and analytic outside of some compact set \( Q_f \subset M \), bounded in modulus by unity and tend to zero at infinity. Clearly \( \alpha(M) \leq \gamma(M) \) and it can be shown that if \( M \) is open then \( \alpha(M) = \gamma(M) \). The analytic capacity of a disk is equal to its radius.
**Theorem.** In order that an arbitrary function which is continuous on the compact set $X$ and analytic on $\text{int} \ X$ be the uniform limit of a sequence of rational functions with poles outside of $X$, it is necessary and sufficient that

$$\alpha(D \setminus X) = \alpha(D \setminus \text{int} \ X)$$

for any open disk $D$.

This condition means that the complement of $X$ is sufficiently massive close to the boundary $\partial X$. It is satisfied in particular if the complement of $X$ is the union of a finite set of finitely connected regions (theorem of Mergelyan). It is also satisfied if $X$ has zero Lebesgue measure so that in this case any function which is continuous on $X$ is the uniform limit of a sequence of rational functions with poles outside of $X$ (theorem of Hartogs and Rosenthal). An example of a compact set on which there is no rational approximation was constructed by E.P. Dolzhenko (1962).

If $X$ is nowhere dense the criterion (76) takes the form

$$\gamma(D \setminus X) = \gamma(D).$$

This is a necessary and sufficient condition for each function which is continuous on $X$ (int $X = \emptyset$) to be the uniform limit of a sequence of rational functions with poles outside of $X$.

Let $X$ be a compact set whose complement $Y$ is simply connected. Let us consider the conformal mapping $w = \psi(z)$ of the domain $Y$ onto the exterior of the unit disk, normalised so that $\psi'(\infty) = 1$. The polynomial part $F_n(z)$ of the Laurent series of the function $[\psi(z)]^n$ ($n = 0, 1, 2, \ldots$) at infinity is called the $n$th Faber polynomial for the compact set $X$. For the closed unit disk this is $z^n$ and for the interval $[-1, 1]$ it is the Chebyshev polynomial $T_n(z)$.

**Theorem of Faber.** Any function $f$ which is analytic on some neighbourhood $U$ of the compact set $X$ can be expanded in terms of Faber polynomials as a series which converges uniformly on some neighbourhood $V$ ($X \subset V \subset U$):

$$f(z) = \sum_{n=0}^{\infty} c_n F_n(z),$$

where (necessarily)

$$c_n = \frac{1}{2\pi i} \int_{|\zeta| = 1 + \varepsilon} f(\psi(\zeta)) \frac{d\zeta}{\zeta^{n+1}} \quad (n = 0, 1, 2, \ldots),$$

(77)

$\psi$ being the inverse function of $\phi$ and $\varepsilon > 0$ sufficiently small.

We can say that this is a conformally invariant form of the Taylor expansion.

Formulae (77) result from the fact that the systems $(F_n(\psi(\zeta)))_0^\infty$ and $(\zeta^{-n-1})_0^\infty$ are biorthogonal on the circle $|\zeta| = 1 + \varepsilon$. Now suppose that we are given any system of analytic functions of the form

$$p_n(z) = \frac{1}{z^{n+1}} + \sum_{k=n+2}^{\infty} a_{nk} \frac{1}{z^k} \quad (|z| > r > 0; n = 0, 1, 2, \ldots).$$
Then there exists a unique sequence of polynomials \((\Phi_n(z))^{\infty}_{n=0}\) \((\deg \Phi_n = n)\) which is biorthogonal with \((p_n(z))^{\infty}_{n=0}\) in the sense that
\[
\frac{1}{2\pi i} \int_{|\zeta|=\rho} \Phi_n(\zeta) p_m(\zeta) d\zeta = \delta_{nm} \quad (\rho > r).
\]
But then we can associate with each function \(f(z)\) which is analytic on the disk \(|z| < R\) \((R > r)\) the series
\[
\sum_{n=0}^{\infty} c_n \Phi_n(z)
\]
with coefficients
\[
c_n = \frac{1}{2\pi i} \int_{|\zeta|=\rho} f(\zeta) p_n(\zeta) d\zeta.
\]
The problem of convergence arises once again in this context, also the moment problem etc.

In particular, if we are given a sequence of interpolating points \(\{z_n\}^{\infty}_{n=0}\) lying in the disk \(|z| < r\), then on putting
\[
P_n(z) = \prod_{k=0}^{n-1} (z - z_k)^{-1}
\]
we will have
\[
\Phi_n(z) = \prod_{k=0}^{n-1} (z - z_k) \quad (n = 0, 1, 2, \ldots)
\]
and (78) becomes Newton's interpolation series. The corresponding moment problem in this case is equivalent to the interpolation problem
\[
f(z_n) = w_n \quad (n = 0, 1, 2, \ldots),
\]
where
\[
w_n = \sum_{k=0}^{n} a_k \prod_{j=0}^{k-1} (z_n - z_j).
\]
We note that the problem (79) can be transformed back into a moment problem of the simpler form
\[
w_n = \frac{1}{2\pi i} \int_{|\zeta|=\rho} \frac{f(\zeta)}{z_n - \zeta} d\zeta.
\]
The uniqueness of the solution of problem (79) is clear since the set of interpolating points has a limit point in the disk \(|z| \leq r < R\). The question of conditions for solvability of this problem is complicated even in the case where the set of limit points for the interpolating points is finite.

In a more subtle version of the problem of type (79) we require that the function \(f\) belongs to the class of analytic functions on the disk \(|z| \leq r\) and all limit points of the set of interpolating points lie on the boundary \(|z| = r\). Such a solution (if
it exists) need not be unique but if, for example, we require in addition that the
function \( f \) is bounded (and for notational simplicity we take \( r = 1 \)) then the
solution is unique if and only if

\[
\lim_{n \to \infty} (1 - |z_n|) = -\infty
\]  

(\textit{theorem of Blaschke}). The lack of uniqueness when (80) fails is shown by the
straightforward construction of the \textit{Blaschke product}:

\[
B(z) = \prod_{n=0}^{\infty} \frac{z_n - z}{1 - \overline{z_n}z} e^{-i\arg z_n}.
\]

Clearly \( |B(z)| < 1 \) for all \( z \) such that \( |z| < 1 \).

The interpolation problem of the form (79) on the disk \( |z| < 1 \) with the
additional boundedness requirement \( |f(z)| < 1 \) is called the \textit{Nevanlinna-Pick problem}\(^{13}\). Positivity of the matrix

\[
\begin{pmatrix}
1 - w_j w_k \\
1 - z_j z_k
\end{pmatrix}_{j,k=0}^{\infty}
\]

is a criterion for its solvability.

It is natural to consider in the class \( E \) of all entire functions, or some subclass
of it, the interpolation problem of the form (79) for a sequence \( z_n \to \infty \). We do
not get uniqueness in \( E \) since there always exists an entire function \( \omega \neq 0 \) with
roots at the points\(^{14}\) \( z_n \) (\( n = 0, 1, 2, \ldots \)) (\textit{theorem of Weierstrass}), although
uniqueness can be restored with 'proper' correlations between the asymptotic
distribution of the points \( z_k \) and the \textit{a priori} restrictions of growth of the
function \( f \).

\textbf{Theorem.} For any sequence of interpolating points \( z_n \to \infty \) and any sequence
of values \( (w_n)_{0}^{\infty} \) there is an entire function \( f \) which satisfies conditions (79).

The general form of such functions is \( f = f_0 + \omega g \), where \( f_0 \) is one of them, \( \omega \)
is an entire function with simple roots \( z_n \) (\( n = 0, 1, 2, \ldots \)) and no other roots and
\( g \) runs through the class \( E \).

For fixed \( \omega \) each sequence \( \{w_n\}_{0}^{\infty} \) defines a \textit{Lagrange interpolation series} (L.i.s):

\[
\sum_{n=0}^{\infty} w_n \omega(z) \omega'(z_n)(z_n - z),
\]

which formally solves problem (79). The investigation of convergence of the L.i.s.
requires powerful analytic tools although the problem is simplified substantially
for a particular choice of interpolating points and class of interpolating functions.

\(^{13}\)A more general version of the Nevanlinna-Pick problem consists of interpolating in a given disk
or half-plane by means of analytic functions with values from another disk or half-plane.

\(^{14}\)It is possible to construct the function \( \omega \) in such a way that all the roots \( z_n \) are simple and there
are no other roots.
Example. Let \( f \) be an entire function of exponential type \( \pi \) which belongs to \( L^2 \) on the real axis. By the Wiener-Paley theorem

\[
\int f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \phi(x)e^{-ix} \, dx,
\]

where \( \phi \in L^2(-\pi, \pi) \) and consequently

\[
\phi(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}
\]

in mean square, where

\[
c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(x)e^{-inx} \, dx = \frac{1}{\sqrt{2\pi}} f(n).
\]

Substituting (82) in (81) we obtain

\[
f(z) = \sum_{n=-\infty}^{\infty} f(n) \frac{(-1)^n \sin \pi z}{\pi(z-n)},
\]

which is an expansion of \( f \) in the L.i.s. corresponding to integer interpolating points. It is easy to see that the series (83) converges uniformly on each compact \( X \subset \mathbb{C} \). This theorem of Kotel'nikov has important applications in the theory of radio communication. It establishes an isometric isomorphism of the space of entire functions \( B_2 \) and the sequence space \( l^2 \). From the physical point of view it effects continuously a discretisation of the transmitted signal.

3.7. Quasianalytic Classes. As Borel showed (1895), for any sequence of numbers \( (a_n)_{\infty} \) there exists an infinitely differentiable function \( f \) on \([0, 1]\) which satisfies the conditions

\[
f^{(n)}(0) = a_n \quad (n = 0, 1, 2, \ldots).
\]

The solution of this 'interpolation' problem is not unique since there exist non-zero functions which are flat (at zero). However if the derivatives of a function \( \phi \in C^\infty[0, 1] \) are subject to bounds of the form

\[
\max_{0 \leq x \leq 1} |\phi^{(n)}(x)| \leq M_n r^n \quad (n = 0, 1, 2, \ldots),
\]

(84)

where \( r > 0 \) and \( M_n > 0 \) is a sequence which does not increase too rapidly, then \( \phi \) cannot be flat if it is non-zero. For example, with \( M_n = n! \) it follows from (84) that \( \phi \) can be extended analytically to the disk \( |z| < r^{-1} \) and therefore if \( \phi \) is flat then \( \phi = 0 \).

Let \( (M_n)_{\infty} \) be any sequence of positive numbers. We say that the function \( \phi \in C^\infty[0, 1] \) belongs to the class \( C(M_n) \) if there exists \( r > 0 \) for which the bounds (84) are satisfied. The class \( C(n!) \) coincides with the class of analytic functions on \([0, 1]\). The class \( C(M_n) \) is said to be quasianalytic if it does not contain flat functions \( \phi \neq 0 \), i.e. if the corresponding uniqueness theorem holds in \( C(M_n) \). The problem of quasianalyticity, which was formulated by Hadamard in 1912, con-
sists of characterising those sequences \((M_n)_{n=1}^\infty\) for which the class \(C(M_n)\) is qua-

sianalytic. Its complete solution was obtained by Carleman (1926) and in rather
different terms by Ostrowski (1930).

**Carleman’s Criterion.** Put

\[
\mu_n = \inf_{k \geq n} \frac{k}{M_k} \quad (n = 1, 2, 3, \ldots).
\]

In order that the class \(C(M_n)\) be quasianalytic it is necessary and sufficient that

\[
\sum_{n=1}^\infty \frac{1}{\mu_n} = \infty.
\]  \hspace{1cm} (85)

**Ostrowski’s Criterion.** Put

\[
T(r) = \sup_{n \geq 0} \frac{r^n}{M_n} (r > 0).
\]

In order that the class \(C(M_n)\) be quasianalytic it is necessary and sufficient that

\[
\int_0^\infty \ln T(r) \frac{dr}{1 + r^2} = \infty.
\]  \hspace{1cm} (86)

The equivalence of conditions (85) and (86) can be demonstrated directly
without reference to the problem of quasianalyticity. We outline the proof of the
sufficiency of condition (86). Without loss of generality we can assume that the
function \(\phi\) is infinitely differentiable on the semiaxis \(x > 0\), satisfies the bounds

\[
|\phi^{(n)}(x)| \leq M_n \quad (x \geq 0; n = 0, 1, 2, \ldots),
\]

belongs to \(L^2(0, \infty)\) and \(\phi^{(n)}(0) = 0\) \((n = 0, 1, 2, \ldots)\). Then its Fourier-Plancherel
transform \(E(\lambda)\) belongs to the class \(H^2\) on the lower half-plane and satisfies the bounds

\[
|E(\lambda)| \leq \frac{M_n}{|\lambda|^n} \quad (n = 1, 2, \ldots),
\]

from which we obtain that on the real axis \(|E_\phi(\lambda)| \leq (T(\lambda))^{-1}\) and, consequently,

\[
\int_{-\infty}^\infty \frac{\ln |E_\phi(\lambda)|}{1 + \lambda^2} d\lambda = -\infty;
\]

this contradicts a well-known property of functions of class \(H^2\).

The necessity of condition (86) can be established by using the construction of
the function \(E(\lambda)\) of class \(H^2\) for a given \(|E_\phi(\lambda)| (-\infty < \lambda < \infty)\).

**§ 4. Integral Equations**

**4.1. Green’s Function.** This object arises naturally in the inversion of linear
differential operators. Let us consider the operator \(L\) on \(C[a, b]\) which is defined
by the expression
for functions \( y \in \mathcal{C}[a, b] \) satisfying conditions of the form

\[
F_i[y] = \int_a^b \left( \sum_{k=0}^{n-1} \alpha_k(x)y^{(k)}(x) \right) d\tau_i(x) = 0 \quad (1 \leq i \leq n).
\]  

(87)

The coefficients \( p_i(x), \alpha_k(x) \) are assumed to be at least continuous and the \( \tau_i(x) \) are functions of bounded variation. If the measures \( \tau_i \) are concentrated at the ends of the interval, then conditions (87) become boundary conditions; in the general case they may be called distributed conditions.

The general solution of the equation \( l[y] = f(f \in \mathcal{C}[a, b]) \) may be written in the form

\[
y = y_0 + \sum_{j=1}^n A_j u_j,
\]

(88)

where \( y_0 \) is a fixed particular solution, \((u_j)_n^1 \) is a basis of the space of solutions of the homogeneous equation \( l[u] = 0 \) and \((A_j)_n^1 \) runs through the space \( \mathbb{C}^n \) (or \( \mathbb{R}^n \) in the real case). Substituting (88) in (87) we obtain the system of linear algebraic equations

\[
\sum_{j=1}^n F_i[u_j] A_j = -F_i[y_0] \quad (1 \leq i \leq n).
\]

(89)

Let us assume that under conditions (87) the homogeneous equation \( l[u] = 0 \) has only the trivial solution. This is equivalent to the homogeneous system

\[
\sum_{j=1}^n F_i[u_j] B_j = 0 \quad (1 \leq i \leq n)
\]

having only the trivial solution. But then the system (89) has a unique solution for any \( y_0 \), i.e. the equation \( L[y] = f \) has a unique solution for any right hand side \( f \in \mathcal{C}[a, b] \). For the explicit construction of the inverse operator \( L^{-1} \) we have to find a particular solution by the method of variation of parameters. As a result it turns out that the operator \( L^{-1} \) is integral,

\[
(L^{-1}f)(x) = \int_a^b G(x, \xi)f(\xi) \, d\xi
\]

(90)

with a certain kernel \( G \). This kernel is called the Green's function of the operator \( L \) and has the following structure

\[
G(x, \xi) = \begin{cases}
\sum_{j=1}^n u_j(x)v_j^{-1}(\xi) & (x \leq \xi), \\
\sum_{j=1}^n u_j(x)v_j^{+1}(\xi) & (x \geq \xi).
\end{cases}
\]

(91)
We can find the functions $v_{j}^{(\pm)}(\xi)$ independently of the calculations mentioned above from the following conditions, which uniquely determine the Green\textsuperscript{Fr} function: for each fixed $\xi (a < \xi < b)$ the function $G(x, \xi)$ belongs to the class $C^{n-2}[a, b]$, its $(n-1)$th derivative exists and is continuous for $x \neq \xi$ while for $x = \xi$ it has a unit jump and the $n$th derivative exists and is continuous for $x \neq \xi$; moreover the differential equation

$$l_{x}[G(x, \xi)] = 0 \quad (x \neq \xi)$$

is satisfied along with the conditions

$$(F_{i})_{x}[G(x, \xi)] = 0 \quad (1 \leq i \leq n).$$

The description which has just been given can be substantially reduced if we invoke the terminology of the $\delta$-function: for each fixed $\xi$ the Green\textsuperscript{Fr} function $G(x, \xi)$ is the solution of the differential equation

$$l_{x}[G(x, \xi)] = \delta(x - \xi)$$

which satisfies the conditions (92). From this point of view formula (90) is obvious since formally

$$l\left[\int_{a}^{b} G(x, \xi)f(\xi) \, d\xi\right] = \int_{a}^{b} \delta(x - \xi)f(\xi) \, d\xi = f(x)$$

and

$$(F_{i})_{x}\left[\int_{a}^{b} G(x, \xi)f(\xi) \, d\xi\right] = \int_{a}^{b} (F_{i})_{x}[G(x, \xi)]f(\xi) \, d\xi = 0 \quad (1 \leq i \leq n).$$

In the case where the conditions (87) are boundary conditions and the coefficients $p_{k}(x)$ are sufficiently smooth (namely $p_{k} \in C^{n-1}[a, b]$) $L$ has an adjoint differential operator $L^*$ defined by the expression\textsuperscript{2}

$$l^{*}[z] = (-1)^{n}p_{n}(x)z^{(n)} + (-1)^{n-1}p_{1}(x)z^{(n-1)} + \cdots + p_{n}(x)z$$

and certain boundary conditions. These are connected with the resulting integration by parts of Lagrange's formula

$$\int_{a}^{b} \{l[y]z - yl^{*}[z]\} \, dx = B(Y(b), Z(b)) - B(Y(a), Z(a)),$$

where $Y(x) = (y(x), y'(x), \ldots, y^{(n-1)}(x))$, $Z(x)$ is defined similarly and $B$ is a bilinear form of the vectors $Y, Z$. It is easy to see that the form $B$ is non-degenerate. Subjecting the function $y$ to the boundary conditions (87), we extract an $n$-dimensional subspace $R$ of the $2n$-dimensional space of vectors $\tilde{Y} = (Y(a), Y(b))$. Let us consider the subspace $R^{\perp}$ of those vectors $\tilde{Z} = (Z(a), Z(b))$ which are orthogonal to every $\tilde{Y} \in R$ in the sense that

\textsuperscript{2}Here for simplicity the concept of the adjoint operator is interpreted in the real sense, which is the natural one if the functions $p_{k}, a_{m}, \tau_{i}$ are real.
\[ B(Y(b), Z(b)) - B(Y(a), Z(a)) = 0. \]

It is \( n \)-dimensional and can be described by the \( n \) boundary conditions
\[ \Phi_i[z] = 0 \quad (1 \leq i \leq n), \tag{87*} \]
where the \( \Phi_i \) are linearly independent linear forms of \( Z \). The differential operator \( L^* \) given by the expression (93) and conditions (87*) is such that
\[ \int_a^b L[y]z \, dx = \int_a^b yL^*[z] \, dx. \tag{94} \]

Moreover, if \( z \in \mathcal{C}[a, b] \) and the identity
\[ \int_a^b L[y]z \, dx = \int_a^b yw \, dx \]
is satisfied for all \( y \in D(L) \) and some \( w \in C[a, b] \), then \( z \in D(L^*) \) and \( w = L^*[z] \).

The differential operator \( L \) is said to be \textit{self adjoint} if \( L^* = L \). Such an operator necessarily has even order \( n = 2m \) and the form
\[ L[y] = y^{(2m)} + \sum_{i=1}^{m} \left( \pi_i(x)y^{(i-m)} + \cdots + \pi_m(x)y \right). \]

Putting \( y = L^{-1}[f] \) in the identity (94) we obtain
\[ \int_a^b f z \, dx = \int_a^b yL^*[z] \, dx. \]

Therefore if \( L^*[z] = 0 \) the function \( z \) is orthogonal to every continuous function and so must be identically zero. Thus we can invert \( L^* \) as well as \( L \) and after the substitution \( z = (L^*)^{-1}[h] \), identity (94) takes the form
\[ \int_a^b \int_a^b G^*(x, \xi)f(x)h(\xi) \, dx \, d\xi = \int_a^b \int_a^b G(x, \xi)h(x)f(\xi) \, dx \, d\xi, \]
where \( G^* \) is the Green's function of the operator \( L^* \). Consequently,
\[ G^*(x, \xi) = G(\xi, x), \]
and for a self-adjoint operator \( L \) its Green's function is symmetric:
\[ G(x, \xi) = G(\xi, x). \]

It is now clear that the systems \((u_i^{(-)})_n \) and \((v_j^{(+)})_n \) in (91) are bases for the space of solutions of the equation \( L^*[u] = 0 \), the \textit{adjoint equation} for \( L[u] = 0 \).

\textbf{Example.} The Green's function of the operator \( L[y] = y'' - q(x)y \) on the interval \([a, b]\) with boundary conditions
\[ y'(a) - h_1y(a) = 0, \quad y'(b) + h_2y(b) = 0 \]
has the form
\[ G(x, \xi) = \begin{cases} \begin{aligned} & u_1(x)u_2(\xi), & (x \leq \xi), \\ & u_2(x)u_1(\xi), & (x > \xi), \end{aligned} \end{cases} \]
where \( \{u_1, u_2\} \) is a basis for the space of solutions of the equation \( L[u] = 0 \), chosen so that \( u_1 \) satisfies the condition at the point \( a \), \( u_2 \) satisfies the condition at the point \( b \) and the Wronskian of this pair of solutions is equal to 1.

We note that by inverting the operator \( L \) the problem of finding eigenvalues and eigenvectors,

\[
L[w] = \lambda w, \tag{95}
\]

is reduced to the homogeneous linear integral equation

\[
w(x) = \lambda \int_a^b G(x, \xi)w(\xi)\,d\xi
\]

with parameter \( \lambda \) (generally complex). If \( \lambda \) is not an eigenvalue then the inhomogeneous equation

\[
L[y] = \lambda y + f
\]

can be solved for any \( f \). Moreover it is equivalent to the integral equation

\[
y(x) = \lambda \int_a^b G(x, \xi)y(\xi)\,d\xi + g(x),
\]

where \( g = L^{-1}[f] \). Thus the types of linear integral equations which have been introduced are naturally associated with linear differential operators and can be used for the study of the latter (in particular, this applies to the Sturm-Liouville problem). The realisation of this approach requires an independent development of the theory of linear integral equations which was carried out in the works of Fredholm, Hilbert and Schmidt; these played a prominent role in the creation of functional analysis.

4.2. Fredholm and Volterra Equations. The definition of these classes does not require linearity; indeed non-linear integral equations also form an interesting and important subject.

Suppose we are given an interval \( I \subset \mathbb{R} \), a function \( g(x) \) on \( I \) and a function \( K(x, \xi, y) \) on \( I \times I \times \mathbb{R} \). An equation of the form

\[
\int_I K(x, \xi, y(\xi))\,d\xi = g(x) \quad (x \in I)
\]

is called a Fredholm integral equation of the 1st kind. By a Fredholm integral equation of the 2nd kind we mean an equation of the form

\[
y(x) = \int_I K(x, \xi, y(\xi))\,d\xi + g(x) \quad (x \in I).
\]

Now let us put \( I_x = I \cap (-\infty, x] \) \( (x \in I) \). If in the above we change the interval of integration to \( I_x \) we obtain Volterra integral equations of the 1st and 2nd kinds in the respective cases.

A general theory of integral equations of the 1st kind is lacking even in the linear case, although certain particular equations of this type have been success-
fully investigated by special methods. Abel's equation is an example of such (see Section 2.10).

For integral equations of the 2nd kind the existence and uniqueness of solutions under specific conditions has been established by means of the contraction mapping principle. In more complicated cases topological methods connected with fixed point principles can often be successfully applied.

**Theorem.** Let $I = [a, b]$, $g \in C[a, b]$ and suppose that the function $K(x, \xi, y)$ on $I \times I \times \mathbb{R}$ is continuous and satisfies a Lipschitz condition for the variable $y$ with constant $M < (b - a)^{-1}$. Then the Fredholm integral equation

$$y(x) = \int_a^b K(x, \xi, y(\xi)) \, d\xi + g(x) \quad (a \leq x \leq b)$$

has a unique continuous solution.

To reduce the problem to the contraction mapping principle it is sufficient to consider the mapping $T$ on $C[a, b]$ which is defined by the right hand side of the equation. Since

$$|((Ty_1)(x) - (Ty_2)(x)| \leq M \int_a^b |y_1(\xi) - y_2(\xi)| \, d\xi \quad (a \leq x \leq b),$$

we have in the uniform metric

$$d(Ty_1, Ty_2) \leq M(b - a) d(y_1, y_2).$$

Therefore the solution not only exists and is unique but it is also the uniform limit on $[a, b]$ of the iterative process

$$y_{n+1}(x) = \int_a^b K(x, \xi, y_n(\xi)) \, d\xi + g(x) \quad (n = 0, 1, 2, \ldots)$$

for any $y_0 \in C[a, b]$.

**Corollary.** Let $I = [a, b]$, $g \in C[a, b]$ and suppose that the function $K(x, \xi, y)$ on $I \times I \times \mathbb{R}$ is continuous and satisfies a Lipschitz condition for the variable $y$ with some constant $M$. Then the Volterra integral equation

$$y(x) = \int_a^x K(x, \xi, y(\xi)) \, d\xi + g(x) \quad (a \leq x \leq b)$$

has a unique continuous solution on the interval $[a, a + l)$ where $l = \min(b - a, M^{-1})$.

We obtain from this, in particular, the theorem of Picard on the existence in some neighbourhood of zero of the solution of the Cauchy problem

$$\frac{dy}{dx} = f(x, y(x)), \quad y(0) = y_0 \quad (96)$$

and its uniqueness, where $f$ is a continuous function which satisfies a Lipschitz
condition for \( y \). In fact, problem (96) is equivalent to the Volterra integral equation

\[
y(x) = \int_0^x f(\xi, y(\xi)) \, d\xi + y_0.
\]

In what follows we will again consider only linear integral equations.

### 4.3. Fredholm Theory

Let us consider the Fredholm integral equation

\[
y(x) = \int_a^b K(x, \xi) y(\xi) \, d\xi + f(x)
\]

(97)
on the interval \( a \leq x \leq b \), assuming that the kernel \( K \) is continuous on the square \( a \leq x, \xi \leq b \) and \( f \in C[a, b] \). We also look for a solution in \( C[a, b] \) (all functions considered can be complex).

The equation (97) is the continuous analogue of the system of algebraic linear equations

\[
y_i = \sum_{j=1}^n \frac{c_{ij}}{k_{ij}} y_j + f_i \quad (1 \leq i \leq n).
\]

(98)
This analogy is highly significant: the original construction of Fredholm theory was based on passage to the limit from the system (98) to the equation (97) as \( n \to \infty \).

The **Fredholm alternative** is the central result of the theory: either the equation (97) has a solution for any \( f \in C[a, b] \) or the **homogeneous** equation

\[
u(x) = \int_a^b K(x, \xi) u(\xi) \, d\xi \quad (a \leq x \leq b)
\]

(99)
has a non-trivial solution. In addition to this we can assert that

1) the subspace of solutions of equation (99) is finite-dimensional,

2) a necessary and sufficient condition for the solvability of equation (97) with fixed \( f \) is that \( f \) be orthogonal to all solutions of the adjoint homogeneous equation

\[
u(x) = \int_a^b K(\eta, x) v(\eta) \, d\eta \quad (a \leq x \leq b),
\]

(99*)

3) the dimensions of the spaces of solutions of equations (99) and (99*) are equal.\(^3\)

Thus the basic facts of the theory of finite systems of algebraic linear equations hold for Fredholm integral equations of the 2nd kind. When applied to the problem of eigenfunctions,

\[
w(x) = \lambda \int_a^b K(x, \xi) w(\xi) \, d\xi,
\]

\(^3\)If they are equal to \( m \) then 2) reduces to \( m \) orthogonality conditions \((f, v_k) = 0 \ (1 \leq k \leq m)\) where \((v_k)_1^m\) is a basis of the space of solutions of equation (99*).
this yields, in particular, the finite-dimensionality of the corresponding eigen-
spaces. Here the eigenvalues turn out to be the roots of a certain entire function
(Fredholm determinant) from which it follows that, if the set of eigenvalues is
infinite, then it can be expressed as a sequence tending to infinity. This last result
was noted in Section 2.11 for the Sturm-Liouville problem but it can now be
regarded as having been established for the more general problem (95).

4.4. Hilbert-Schmidt Theory. The Fredholm theory remains valid for the space
$L^2(I)$ (even in the case of an infinite interval $I$) if it is assumed that $K(x, \xi)$ is a
Hilbert-Schmidt kernel, i.e. if it is square integrable as a function of $x$ and $\xi$ and
Hermitian: $K(\xi, x) = \overline{K(x, \xi)}$ (these conditions are satisfied if $I$ is finite and $K$
is the Green function of a self adjoint linear differential operator). All the eigen-
values then turn out to be real and eigenfunctions corresponding to distinct
eigenvalues are orthogonal. We have here a complete analogue of the spectral
theory of Hermitian matrices, the only difference being that the set of eigenvalues
is infinite because the space $L^2(I)$ is infinite-dimensional. In fact we associate with
the kernel $K(x, \xi)$ a (Hermitian) quadratic functional

$$Q[y] = \int_I \int_I K(x, \xi) y(x)\overline{y(\xi)} \, dx \, d\xi$$

just as a (Hermitian) quadratic form is associated with any Hermitian matrix.

The variational method applied to the functional $Q$ under the condition

$$\int_I |y(x)|^2 \, dx = 1$$

leads to the following fundamental result of Hilbert-Schmidt theory.

**Theorem.** Suppose that the equation

$$\int_I K(x, \xi) u(\xi) \, d\xi = 0$$

with a Hilbert-Schmidt kernel has only the trivial solution in $L^2(I)$. Then the
eigenvalues form an infinite sequence tending to infinity and a complete ortho-
normal system $(\phi_n)_{n=1}^{\infty}$ can be constructed from its eigenfunctions.

A most important consequence of this result is the theorem on the complete
system of eigenfunctions in the Strum-Liouville problem and, in general, for any
linear self-adjoint differential operator on a finite interval.

Let us denote by $\lambda_n$ the eigenvalue corresponding to the eigenfunction $\phi_n$. Then
we have the expansion of the kernel

$$K(x, \xi) = \sum_{n=1}^{\infty} \frac{\phi_n(x)\overline{\phi_n(\xi)}}{\lambda_n}, \quad (100)$$

4 We can even prove the finite-dimensionality of the collection of root-spaces

$$W_r(\lambda) = \{w: (1 - \lambda K)w = 0\} \quad (r = 1, 2, 3, \ldots)$$

($K$ is the integral operator with kernel $K(x, \xi)$, 1 is the identity operator).
which converges in mean square on $I \times I$. Consequently

$$Q[y] = \sum_{n=1}^{\infty} \frac{|c_n|^2}{\lambda_n},$$

(101)

where the $c_n$ are the Fourier coefficients of the function $y$ with respect to the system $(\phi_n)_1^\infty$, i.e.

$$c_n = \int_I y(x) \overline{\phi_n(x)} \, dx.$$

The identity (101) is just the theorem on the reduction of the Hermitian quadratic functional $Q$ to its principal axes.

We note that the relation

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n^2} = \int_I \int_I |K(x, \xi)|^2 \, dx \, d\xi$$

follows from (100) on applying Parseval's equality. If all the $\lambda_n > 0$ (which is equivalent to positivity of the functional $Q$) and if further the kernel $K$ is continuous and the interval $I$ is finite, then the expansion (100) converges uniformly on the square $I \times I$ (theorem of Mercer). In particular

$$K(x, x) = \sum_{n=1}^{\infty} \frac{|\phi_n(x)|^2}{\lambda_n},$$

from which follows the trace formula:

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \int_I K(x, x) \, dx.$$

4.5. Equations with Difference Kernels. We are concerned with integral equations whose kernels are of the form

$$K(x, \xi) = a(x - \xi) \quad (x, \xi \in I)$$

where the function $a(t)$ (also called the kernel) is defined on the interval $J = I - I$.

First let us consider the equation

$$y(x) = \int_{-\infty}^{\infty} a(x - \xi)y(\xi) \, d\xi + g(x) \quad (-\infty < x < \infty)$$

(102)

on the whole axis and in $L^1(-\infty, \infty)$ (the functions $a, g$ are given in this class and we look for $y$ in it). In terms of convolution, equation (102) has the form

$$y = a \ast y + g.$$

**Theorem 1.** In order that equation (102) have a solution $y \in L^1$ for each $g \in L^1$ it is necessary and sufficient that the Fourier transform

---

5 It is to be understood that (102) must be satisfied only almost everywhere.
be different from 1 for all \( \lambda \).

In fact, using the language of Fourier transforms, we can express equation (102) algebraically:

\[
Y(\lambda) = A(\lambda) Y(\lambda) + G(\lambda).
\]

If \( A(\lambda_0) = 1 \) and the equation is solvable we obtain \( G(\lambda_0) = 0 \), which cannot hold for all \( g \in L^1 \). Conversely, if \( A(\lambda) \neq 1 \) for all \( \lambda \), then by Wiener's theorem

\[
(1 - A(\lambda))^{-1} = 1 + B(\lambda),
\]

where \( B \) is the Fourier transform of a certain function \( b \in L^1 \). Consequently equation (102) is satisfied by

\[
y = g + b \ast g
\]

and clearly this is the unique solution.

The theory of the equation

\[
y(x) = \int_{0}^{\infty} a(x - \xi) y(\xi) \, d\xi + g(x) \quad (x > 0) \tag{103}
\]

on the semiaxis and in the class \( L^1(0, \infty) \) with kernel \( a \in L^1(-\infty, \infty) \) is considerably more complicated. A technique for solving equation (103) in the case where \( g = 0 \) and the kernel decreases exponentially was indicated by Wiener and E. Hopf in 1931 and then generalised and developed by many authors. Moreover deep connections with the theory of analytic functions, functional analysis and even topology were found. The topological aspect is connected with the use of the index

\[
v = \text{ind}(1 - A) \equiv -\frac{1}{2\pi} \arg(1 - A(\lambda))|_{-\infty}^{\infty}
\]

(for solvability of the equation for all \( g \in L^1 \) it is necessary as before that \( A(\lambda) \neq 1 \) for all \( \lambda \) and we will assume that this condition is satisfied in what follows). Because the curve \( \zeta = 1 - A(\lambda) \) \( (-\infty \leq \lambda \leq \infty; A(\pm \infty) = 0) \) is closed the index is always an integer.

**Theorem 2.** For solvability of equation (103) for all \( g \in L^1 \) it is necessary and sufficient that the index be non-negative. For uniqueness of the solution it is necessary and sufficient that the index be equal to zero.

We shall describe a method of proof which goes back to Wiener and Hopf and was sufficiently generalised by M.G. Krejn (1958). It consists of factorisation of the function \( (1 - A(\lambda))^{-1} \), i.e. the decomposition

\[\text{[Footnote]}\]

\[\text{[Footnote]}\]

\( \text{[Footnote]}\) The theory presented below extends to many other classes, in particular to \( L^p \) \( (p > 1) \).
\[(1 - A(\lambda))^{-1} = [1 + B_+(\lambda)][1 + B_-(\lambda)] \quad (-\infty < \lambda < \infty),\]

where

\[B_+(\lambda) = \int_0^\infty b_+(t)e^{-\lambda t} \, dt, \quad B_-(\lambda) = \int_0^\infty b_-(t)e^{-\lambda t} \, dt\]

are functions which can be extended analytically to the upper and lower half-planes respectively \((b_+ \in L^1(0, \infty))\). If \(v \geq 0\) the factor \(1 + B_-(\lambda)\) can always be chosen so that it has no roots in the lower half-plane. Moreover

\[(1 + B_-(\lambda))^{-1} = 1 + \int_0^\infty \beta_-(t)e^{-\lambda t} \, dt \quad (\text{im } \lambda \leq 0),\]

where \(\beta_- \in L^1(0, \infty)\). The factor \(1 + B_+(\lambda)\) will have exactly \(v\) roots (counted according to multiplicity) in the upper half-plane (by the principle of the argument); their locations can be chosen arbitrarily, after which the factorisation will be uniquely determined (so that for \(v = 0\) it is unique). More specifically

\[1 + B_+(\lambda) = R(1)(1 + c(\lambda)),\]

where \(R(\lambda)\) is a rational function which has the prescribed roots in the half-plane \(\text{im } \lambda > 0\), a pole of order \(v\) at \(\lambda = -i\) and no other roots or poles, and

\[(1 + c(\lambda))^{-1} = 1 + \int_0^\infty \beta_+(t)e^{\lambda t} \, dt \quad (\text{im } \lambda \geq 0),\]

where \(\beta_+ \in L^1(0, \infty)\). For \(v = 0\)

\[(1 + B_+(\lambda))^{-1} = 1 + \int_0^\infty \beta_+(t)e^{\lambda t} \, dt \quad (\text{im } \lambda \geq 0).\]

For \(v = 0\) the factorisation is brought about by the possibility of choosing a unique branch of \(\ln(1 - A(\lambda))\) which tends to zero as \(\lambda \to \pm \infty\). It turns out that

\[\ln(1 - A(\lambda)) = \int_{-\infty}^\infty l(t)e^{\lambda t} \, dt \quad (-\infty < \lambda < \infty)\]

(a special case of a theorem of Wiener and Levy), hence

\[\frac{1}{1 - A(\lambda)} = \exp\left[-\int_{-\infty}^0 l(t)e^{\lambda t} \, dt\right] \exp\left[-\int_0^\infty l(t)e^{\lambda t} \, dt\right]\]

and further the factors obtained can be written in the form

\[1 + \int_0^\infty b_+(t)e^{\lambda t} \, dt, \quad 1 + \int_0^\infty b_-(t)e^{\lambda t} \, dt\]

(also with the help of the same theorem of Wiener and Levy).

For \(v > 0\) the index is cancelled out by the introduction of the factor \(R(\lambda)\).

To solve equation (103) when \(v = 0\) we extend the function \(g\) and the unknown function \(y\) by setting them equal to zero on the semiaxis \(x < 0\), after which we
can write the equation in the form

$$y(x) = \int_{-\infty}^{\infty} a(x - \xi) y(\xi) \, d\xi + g(x) + h(x) \quad (-\infty < x < \infty),$$

where

$$h(x) = -\int_{0}^{\infty} a(x - \xi) y(\xi) \, d\xi \quad (x < 0)$$

and $h(x) = 0 \ (x \geq 0)$. Converting to Fourier transforms and using the factorisation we obtain

$$Y(\lambda) \frac{1}{1 + B_-(\lambda)} = (1 + B_+(\lambda))G(\lambda) + (1 + B_+(\lambda))H(\lambda). \quad (104)$$

The left hand side of this equality is the Fourier transform of some function from $L^1(0, \infty)$ extended by zero on the left. On the right hand side the first term is the Fourier transform of the function $f = g + \beta \ast g$, where $\beta(x) = b_+(-x) \ (x < 0)$ and $\beta(x) = 0 \ if \ x \geq 0$. The second term is the Fourier transform of a certain function from $L^1(-\infty, 0)$ extended by zero on the right.

Consequently

$$Y(\lambda) = (1 + B_-(\lambda)) \int_{0}^{\infty} f(x)e^{-i\lambda x} \, dx. \quad (105)$$

The arguments outlined\(^7\) can be reversed, i.e. (105) is in fact the solution (it is unique because of the previous part).

For $\nu > 0$ each factorisation produces a certain solution from formula (105). Thus we obtain an infinite set of solutions parameterised by divisors of roots with total multiplicity $\nu$ which lie in the upper half-plane. The space $N$ of solutions of the homogeneous equation

$$u(x) = \int_{0}^{\infty} a(x - \xi) u(\xi) \, d\xi \quad (x > 0) \quad (106)$$

is $\nu$-dimensional in the given case. All the functions appearing in $N$ are bounded\(^8\). For $\nu < 0$ the adjoint\(^9\) homogeneous equation

$$v(x) = \int_{0}^{\infty} a(\xi - x) v(\xi) \, d\xi \quad (x > 0) \quad (106^*)$$

has index $(-\nu) > 0$ and so the space $N'$ of its solutions is $|\nu|$-dimensional. If equation (103) is solvable in this case, $\vec{g}$ must be orthogonal to all solutions of equation (106\(^*)\) (and, conversely, under this condition equation (103) is solvable). Thus a necessary condition for solvability of equation (103) for all $g$ is that $\nu \geq 0$.

\(^7\)We draw attention to their similarity with separation of variables.

\(^8\)A suitable construction of a basis in $N$ shows that each function $u \in N$ is absolutely continuous, $u' \in L^1$ and $u(\infty) = 0$.

\(^9\)More precisely, transposed, but we will not make use of this term, relying upon the context.
This situation is related to the Fredholm case but in general there is the significant difference that \( \dim N' \neq \dim N \). The difference \( \dim N - \dim N' \) is equal to the index \( v \) in all cases, i.e.

\[
\dim N = \begin{cases} 
v & (v \geq 0) \\
0 & (v < 0) \end{cases}, \quad \dim N' = \begin{cases} 
0 & (v \geq 0) \\
-v & (v < 0) \end{cases}.
\]

Such a situation is typical of singular integral equations, i.e. equations whose kernels have sufficiently strong singularities. The first general results in the theory of singular integral equations were obtained by M. Noether (1921). This theory is closely connected with the so-called Riemann-Hilbert problem, a particular case of which is the relation (104). Its solution was given by F.D. Gakhov (1936).

4.6. The Riemann-Hilbert Problem. Let \( \Gamma \) be a simple closed smooth contour in the extended complex plane \( \mathbb{C} = \mathbb{C} \cup \{ \infty \} \) and let \( a(\zeta), b(\zeta) \) be two continuous functions on \( \Gamma \) with \( a(\zeta) \neq 0 \) for all \( \zeta \in \Gamma \). The complement \( \mathbb{C} \setminus \Gamma \) consists of two domains \( G_i \) and \( G_e \) (interior and exterior; \( \infty \in G_e \cup \Gamma \)). The problem is to find two functions \( f_i(\zeta) (\zeta \in G_i) \), \( f_e(\zeta) (\zeta \in G_e) \) which are analytic on the respective domains\(^{10} \), continuous up to the boundary and connected by the relation

\[
f_i(\zeta) = a(\zeta)f_e(\zeta) + b(\zeta) \quad (\zeta \in \Gamma).
\]

The index of this problem is defined as

\[
v = \frac{1}{2\pi} \Delta_\Gamma \arg a(\zeta),
\]

where \( \Delta_\Gamma \) is the increment for a circuit of the contour \( \Gamma \) with \( G_i \) on the left.

In (104) \( \Gamma = \mathbb{R} = \mathbb{R} \cup \{ \infty \} \), \( G_i \) is the lower half-plane, \( G_e \) is the upper half-plane, \( a(\zeta) = (1 - A(\zeta))^{-1} \), \( b(\zeta) = (1 - A(\zeta))^{-1} G(\zeta) \), \( f_i(\zeta) = Y(\zeta) \) and \( f_e(\zeta) = H(\zeta) \). Usually a bounded contour \( \Gamma \) is considered and it is assumed that \( a, b \in \text{Lip} \delta \) for some \( \delta \). Let us suppose further that these conditions are satisfied and assume for convenience that \( 0 \in G_i \).

If \( v = 0 \) there is defined on the contour \( \Gamma \) a single-valued function \( \alpha(\zeta) = \ln a(\zeta) \), which allows us to take the logarithm in the corresponding homogeneous problem

\[
g_i(\zeta) = a(\zeta)g_e(\zeta) \quad (\zeta \in \Gamma).
\]

The additive problem

\[
\phi_i(\zeta) - \phi_e(\zeta) = \alpha(\zeta) \quad (\zeta \in \Gamma)
\]

now arises, the solution of which (unique if \( \phi_e(\infty) = 0 \)) is given by the integral of Cauchy type

\[
\frac{1}{2\pi} \int_{\Gamma} \frac{\alpha(\zeta) d\zeta}{\zeta - z} = \begin{cases} 
\phi_i(z) & (z \in G_i), \\
\phi_e(z) & (z \in G_e).
\end{cases}
\]

\(^{10}\)If \( \infty \in G_e \) then \( f_e \) has to be regular at infinity.
This produces for \( a(\zeta) \) the factorisation

\[
a(\zeta) = \frac{g_i(\zeta)}{g_e(\zeta)} \quad (\zeta \in \Gamma),
\]

where \( g_i, g_e \) are the exponentials of \( \phi_i, \phi_e \). They are continuous on \( \overline{G_i}, \overline{G_e} \), analytic on \( G_i, G_e \) respectively and have no roots. Such a factorisation is defined up to a common constant factor in \( g_i \) and \( g_e \).

If \( \nu > 0 \) we can nullify the index by dividing \( a(\zeta) \) by \( \zeta^\nu \), after which we can apply the previous factorisation. This produces the desired factorisation of \( a(\zeta) \); moreover \( g_e(z) = O(|z|^{-\nu}) \) as \( |z| \to \infty \) which leaves the possibility of multiplying the solution obtained, \((g_i, g_e)\), by an arbitrary polynomial of degree \( \leq v \). All solutions are realised by this process, so that the space of solutions turns out to be \((v + 1)\)-dimensional\(^1\).

Just as in the Wiener-Hopf method, using the factorisation we can effectively solve the inhomogeneous problem by the separation of variables. It turns out to be solvable for all \( b(\zeta) \) if \( \nu \geq -1 \). A criterion for solvability when \( \nu < -1 \) is the orthogonality condition

\[
\int_{\Gamma} b(\zeta) h_i(\zeta) \, d\zeta = 0,
\]

where \( h_i \) is the interior part of the general solution of the associated homogeneous problem

\[
h_i(\zeta) = \frac{1}{\zeta^2 a(\zeta)} h_e(\zeta) \quad (108^*)
\]

(its index is equal to \(-(v + 2))\).

If for any index we denote by \( N \) the space of solutions of problem \((108)\) and by \( N' \) the space of solutions of problem \((108^*)\) (with the previous requirement of regularity at infinity), then

\[
\dim N = \begin{cases} 
  v + 1 & (\nu \geq 0), \\
  0 & (\nu < 0)
\end{cases} \quad \dim N' = \begin{cases} 
  0 & (\nu \geq -1), \\
  -\nu - 1 & (\nu < -1)
\end{cases}.
\]

Therefore in all cases

\[
\dim N - \dim N' = v + 1. \quad (109)
\]

For \( \nu = -1 \) the situation becomes that of the Fredholm case.

The Riemann-Hilbert problem reduces to a singular integral equation if we look for a solution in the form of an integral of Cauchy type

\[
\frac{1}{2\pi i} \int_{\Gamma} \frac{\psi(\zeta) \, d\zeta}{\zeta - z} = \begin{cases} 
  f_i(z) & (z \in G_i) \\
  \hat{f}_e(z) & (z \in G_e)
\end{cases} \quad (108^*)
\]

(in addition the condition \( \hat{f}_e(\infty) = 0 \) arises). By Sokhotskij's formula (with the

\(^1\) For \( \nu = 0 \) it was one-dimensional.
assumption that $\psi \in \text{Lip } \delta$),

$$f_{i, \epsilon}(\zeta) = \pm \frac{1}{2} \psi(\zeta) + \frac{1}{2\pi i} \int_\Gamma \frac{\psi(\tau) d\tau}{\tau - \zeta} \quad (\zeta \in \Gamma),$$

where the integral is to be interpreted as a principal value, i.e. as the limit of the integral along the arc $\Gamma_\epsilon(\zeta) = \{\tau: \tau \in \Gamma, |\tau - \zeta| > \epsilon\}$ as $\epsilon \to 0$. Substitution in (107) leads to a singular equation of the form

$$\alpha(\zeta)\psi(\zeta) + \beta(\zeta) \int_\Gamma \frac{\psi(\tau) d\tau}{\tau - \zeta} = \gamma(\zeta) \quad (\zeta \in \Gamma), \quad (110)$$

where $\alpha(\zeta) = (1 + a(\zeta))/2$, $\beta(\zeta) = (1 - a(\zeta))/2$, $\gamma(\zeta) = b(\zeta)$ and consequently $\alpha(\zeta) + \beta(\zeta) \equiv 1$.

Conversely, if an equation of the form (110) is given and $\alpha(\zeta) + \beta(\zeta) \neq 0$ everywhere on $\Gamma$, then (by inversion of the above calculations) it reduces to the Riemann-Hilbert problem with coefficients

$$a(\zeta) = \frac{\alpha(\zeta) - \beta(\zeta)}{\alpha(\zeta) + \beta(\zeta)}, \quad b(\zeta) = \frac{\gamma(\zeta)}{\alpha(\zeta) + \beta(\zeta)},$$

which allows us to define the index of equation (110) and to obtain the same facts that are available for the Riemann-Hilbert problem. These constitute the theory of Noether:

1) finite-dimensionality of the space $N$ of solutions of the homogeneous equation;

2) a criterion for solvability of the inhomogeneous equation in terms of orthogonality of the right hand side to the space $N'$ of solutions of the adjoint homogeneous equation;

3) the equality $\dim N - \dim N' = v$ (the difference from (109) is explained by the fact that $f_\epsilon(\infty) = 0$).

The theory of Noether remains valid for more general singular equations

$$\alpha(\zeta)\psi(\zeta) + \frac{\beta(\zeta)}{\pi i} \int_\Gamma \frac{\psi(\tau) d\tau}{\tau - \zeta} + \int_\Gamma k(\zeta, \tau)\psi(\tau) d\tau = \gamma(\zeta) \quad (\zeta \in \Gamma),$$

where the kernel $k(\zeta, \tau)$ has a 'weak' singularity:

$$k(\zeta, \tau) = O(|\zeta - \tau|^{-\kappa}), \quad 0 < \kappa < 1.$$
Chapter 2
Foundations and Methods

§ 1. Infinite-Dimensional Linear Algebra

All the situations discussed in this paragraph will be concerned with a prescribed basic linear space $E \neq 0$ over an arbitrary (within the limits of this section) field $K$ (or simultaneously with several such spaces). The elements of the space $E$, i.e. vectors or points, will usually be denoted by lower case Roman letters, while lower case Greek letters will be used for scalars, i.e., elements of the ground field. The identity of the field will be denoted by the figure 1.

The standard function spaces ($C, L^p$ etc.) lie outside of the range of traditional linear algebra, since it is restricted to the finite-dimensional case. Consequently in the context of functional analysis this restriction is unacceptable, the reason being that the spaces considered below are generally infinite-dimensional.

Infinite-dimensional linear algebra comes in at the foundation of functional analysis as a suitable 'algebraic model' for the theory of linear topological spaces, which occupies a central place in functional analysis.

1.1. Bases and Dimension. By an algebraic\(^1\) basis, or Hamel basis, of a space $E$ we mean a maximal (with respect to inclusion) linearly independent subset. Using Zorn’s lemma (or, equivalently, transfinite induction) we can easily prove

**Theorem 1.** Any linearly independent subset of a space $E$ can be extended to a basis.

**Corollary 1.** There exists a basis in $E$.

**Corollary 2.** A basis of any subspace $L \neq 0$ can be extended to a basis of the whole space $E$.

Moreover the extending vectors\(^2\) form a basis of a subspace $M$ such that the whole space $E$ is the direct sum of $L$ and $M$ ($L + M = E$, i.e. $M$ is a complement for $L$).

Let a basis $\{e_i\}_{i \in I}$ of the space $E$ be given ($I$ is an index set). In the unique representation of a vector $x \in E$ with respect to the given basis the set of indices of coefficients which are different from zero is finite. It is called the support of the vector $x$ and is denoted by $\text{supp } x$.

We have

$$x = \sum_{i \in \text{supp } x} \xi_i e_i,$$

\(^1\)This word can be omitted in the purely algebraic context of the present paragraph.

\(^2\)If they exist.
which can also be written formally as

\[ x = \sum_{i \in I} \xi_i e_i \]

(the \( \xi_i \) are called the coordinates of the vector \( x \) with respect to the given basis).

Any other basis \( \{u_j\}_{j \in J} \) can be expressed in the form

\[ u_j = \sum_{i \in I} \alpha_{ij} e_i, \]

where the transfer matrix \( \alpha = (\alpha_{ij}) \) is infinite (if \( E \) is infinite-dimensional) but each of its columns has finite support, i.e., it contains only a finite set of non-zero elements; in addition, the matrix \( \alpha \) is invertible: if

\[ e_k = \sum_{j \in J} \beta_{kj} u_j \quad (k \in I) \]

and \( b = (\beta_{jk}) \), then

\[ \sum_{j \in J} \alpha_{ij} \beta_{jk} = \delta_{ik} \quad (i, k \in I), \quad \sum_{i \in I} \beta_{ji} \alpha_{il} = \delta_{jl} \quad (j, l \in J). \]  

(1)

If \( x \) has coordinates \( (\xi_i)_{i \in I} \) with respect to the initial basis and \( (\eta_j)_{j \in J} \) with respect to the new basis, then

\[ \xi_i = \sum_{j \in J} \alpha_{ij} \eta_j. \]

The formal distinction between the index sets \( I \) and \( J \) is not essential since we have

**Theorem 2.** All bases of the space \( E \) are equipotent.

**Proof.** In the previous notation we have the mapping

\[ i \mapsto \text{supp } e_i \quad (i \in I) \]

of the set \( I \) into the set of finite subsets \( A \) of the set \( J \). The preimage of each \( A \) is finite. Therefore \( \text{card } I \leq \text{card } J \). Similarly, \( \text{card } J \leq \text{card } I \). \( \square \)

The power (cardinality) of any basis is called the algebraic dimension of the space \( E \) and is denoted by \( \text{dim } E \). As a common index set for all bases we can take the least ordinal of power \( \text{dim } E \). Let us denote this by \( n \).

We note that, after identifying \( J \) with \( I \), we can write formulae (1) in the form \( ab = e \), \( ba = e \), where \( e = (\delta_{nk}) \) is the identity matrix. The invertible matrices whose columns have finite support form a full linear group \( GL(n, K) \).

On any set \( S \) a function \( \varphi: S \to K \) is said to have finite support if its support \( \text{supp } \varphi = \{s: \varphi(s) \neq 0\} \) is finite\(^3\). The functions of finite support form a subspace \( \mathcal{F}_0(S, K) \) of the space \( \mathcal{F}(S, K) \) of all functions (it is a proper subspace if the set \( S \) is infinite). In \( \mathcal{F}_0(S, K) \) we have the canonical basis \( \{\delta_s\}_{s \in S}: \delta_s(t) = \delta_{st} (t \in S) \). The

\(^3\)A function \( \varphi \) on a topological space is said to have compact support if its support \( \text{supp } \varphi = \{s: \varphi(S) \neq 0\} \) is compact. This coincides with the present notion of we give \( S \) the discrete topology.
representation of any \( \varphi \in \Phi_0(S, K) \) is
\[
\varphi = \sum_{s \in \text{supp}\varphi} \varphi(s)\delta_s = \sum_{s \in S} \varphi(s)\delta_s.
\]

Since, clearly, \( \dim \Phi_0(S, K) = \text{card } S \), any cardinal is the dimension of some linear space.

The dimensions of isomorphic spaces \( (E_1 \cong E_2) \) are obviously equal and conversely isomorphism of the spaces follows from equality of their dimensions (a 1–1 correspondence between the bases can be extended by linearity to an isomorphism of the spaces).

**Example.** If \( B \) is a basis in \( E \), then \( E \cong \Phi_0(B, K) \). Hence it follows in particular that, if \( K \) is infinite or \( E \) is infinite-dimensional, then
\[
\text{card } E = \max(\dim E, \text{card } K).
\]

We can use this formula to calculate the cardinal \( \nu = \dim \Phi(S, K) \) for infinite \( S \) (if \( S \) is finite then \( \nu = \text{card } S \)). We have
\[
\text{card } \Phi(S, K) = \max(\nu, \text{card } K).
\]

But \( \nu \geq \text{card } K \), for even if \( S = \mathbb{N} = \{0, 1, 2, \ldots\} \) the functions \( \varphi_\alpha(s) = \alpha^s (\alpha \in K) \) are linearly independent. Thus
\[
\dim \Phi(S, K) = \text{card } \Phi(S, K) = (\text{card } K)^{\text{card } S}.
\]

We note that \( \dim \Phi(S, K) > \text{card } S \), i.e.
\[
\dim \Phi(S, K) > \dim \Phi_0(S, K).
\]

An interesting quantity associated with a linear space \( E (\dim E > 1) \) is the least power for a set of proper subspaces covering \( E \). It turns out to be countable if \( E \) is infinite-dimensional and \( K \) is infinite,
equal to \( \text{card } K \) if \( E \) is finite-dimensional and \( K \) is infinite,
equal to \( \text{card } K + 1 \) if \( K \) is finite.

The intersection and sum are defined for any set \( \mathcal{L} = \{L_\sigma\} \) of subspaces. The latter consists of all possible finite sums \( x_{\sigma_1} + \cdots + x_{\sigma_n} (x_{\sigma} \in L_\sigma) \). The set \( \mathcal{L} \) is said to be independent if \( L_\sigma \cap \sum_{\tau \neq \sigma} L_\tau = 0 \) for all \( \sigma \). The sum of an independent set of subspaces is called a direct sum\(^5\) and is denoted by \( \sum \): Its basis can be obtained by combining the bases of the summands and therefore its dimension

\(^2\) If \( K \) is finite and \( E \) is finite-dimensional then clearly
\[
\text{card } E = (\text{card } K)^{\dim E}.
\]

\(^3\) We can also define the external direct sum \( \Sigma E_\sigma \) of a set of spaces \( \{E_\sigma\} \) as the subspace of the cartesian product \( \Pi E_\sigma \) which consists of those elements \( x = (x_\sigma) \) with finite support. Further, if there is a scalar product \( (.,.) \) on each \( E_\sigma \), we can introduce a scalar product in \( \Sigma E_\sigma \) by putting \( (x, y) = \Sigma(x_\sigma, y_\sigma) \). In this case we refer to the orthogonal sum of the spaces \( E_\sigma \) and use the symbol \( \oplus \) in place of \( + \).
is equal to the sum of their dimensions. In the general case the dimension of a sum of subspaces does not exceed the sum of their dimensions.

The intersection of all subspaces which contain a given non-empty set \(X \subset E\) (i.e. the smallest such subspace) coincides with the linear hull \(\text{Lin} \; X\), i.e. with the set of all linear combinations of vectors in \(X\). The dimension \(\dim(\text{Lin} \; X)\) is called the rank of the set \(X\) and is denoted by \(\text{rk} \; X\). It is equal to the power of any maximal linearly independent subset \(Y \subset X\) (each such \(Y\) is a basis in \(\text{Lin} \; X\)).

One of the fundamental constructions of linear algebra associates with each subspace \(L \subset E\) the factor space \(E/L\), consisting of the classes of the following equivalence relation (congruence modulo \(L\)):

\[x \equiv y(\text{mod} \; L) \iff x - y \in L.\]

The class of the vector \(x\) is denoted by \([x]\). The operations on the classes are defined naturally by

\[[x] + [y] = [x + y], \quad \alpha [x] = [\alpha x],\]

which are valid because the right hand sides of these identities are independent of the choice of representatives of the classes on the left hand sides.

Any complement \(M\) of the subspace \(L\) is isomorphic to \(E/L\) (the natural mapping \(x \mapsto [x]\) serves as the isomorphism). All complements of a given subspace \(L\) are therefore isomorphic and consequently they all have the same dimension. It is called the codimension of the subspace \(L\) and is denoted by \(\text{codim} \; L\). If \(\text{codim} \; L = 1\) (and only in this case) \(L\) is a maximal (with respect to inclusion) subspace different from \(E\) and is called a hyperplane.

A subset of the form \(A = x + L\) (\(x \in E, \; L\) a subspace) is called an affine manifold. In addition we put \(\dim A = \dim L\), \(\text{codim} \; A = \text{codim} \; L\). If \(\dim A = 1\) it is called a line; if \(\text{codim} \; A = 1\) it is a hyperplane.

1.2. Homomorphisms and Linear Functionals. For any linear spaces \(E_1, \; E_2\) the linear space of homomorphisms\(^6\) \(h: E_1 \to E_2\) is denoted by \(\text{Hom}(E_1, \; E_2)\). In particular \(\text{Hom}(E, \; K)\) is the space of linear functionals on \(E\). It is called the algebraic conjugate of \(E\) and is denoted by \(E^*\).

If a basis \(B = \{e_i\}_{i \in I}\) is given in \(E\) then to each \(f \in E^*\) there corresponds a function \(f(e_i)\) on \(B\) and this correspondence is an isomorphism (so that \(E^* \approx \Phi(B, \; K)\))\(^7\) since

\[f(x) = \sum_{i \in I} f(e_i) \xi_i(x),\]

where the \(\xi_i(x)\) are the coordinates of the vector \(x\) with respect to the basis \(B\).

We stress that the coordinate functionals \(\xi_i\) do not form a basis in \(E^*\) if \(E\) is infinite-dimensional, since the infinituples \((f(e_i))_{i \in I}\) can be chosen arbitrarily and

---

\(^6\) According to the standard definition, \((h_1 + h_2)x = h_1x + h_2x, (ah)x = a(hx)\). Further, if \(h: E_1 \to E_2, \; g: E_2 \to E_3\) are homomorphisms, then their product (superposition or composition) \(gh: E_1 \to E_3\) is also a homomorphism.

\(^7\) In other words \(\Phi_0(S, \; K)^* \approx \Phi(S, \; K)\) for any \(S\).
not just with finite support. However the set \( \{ \xi_t \}_{t \in T} \) is linearly independent because of the biorthogonality relation: \( \xi_t(e_k) = \delta_{tk} \). As a result of (2), \( \dim E^* > \dim E \) if (and only if) \( E \) is infinite-dimensional.\(^8\)

If \( L \) is a subspace of \( E \), the restriction \( f|_L \) of any \( f \in E^* \) is a linear functional on \( L \). Conversely, any \( f_0 \in L^* \) can be extended to an \( f \in E^* \) by choosing any complement \( M \) for the subspace \( L \) and setting \( f(x) = f_0(u) \) where \( x = u + v \) \((u \in L, v \in M)\). Hence it follows in particular that if \( e \in E \) and \( e \neq 0 \) there exists \( f \in E^* \) such that \( f(e) \neq 0 \) (it is possible to make \( f(e) = 1 \)). The same result shows that the linear functionals on \( E \) separate points: if \( x, y \in E \) and \( x \neq y \) there exists \( f \in E^* \) such that \( f(x) \neq f(y) \).

When applied to \( L = \text{Lin} \{e\} \), the extension method described above shows that, if a subspace \( V \) does not contain the vector \( e \), there exists \( f \in E^* \) such that \( f(e) \neq 0 \) and \( f|_V = 0 \). Any subspace is therefore the set of solutions of a system of linear equations \( f_i(x) = 0 \) \((\{f_i\} \subset E^*)\).

The subspace

\[
L^\perp = \{ f: f \in E^*, f(x) = 0 \ (x \in L) \}
\]

of linear functionals which are orthogonal to the subspace \( L \) is called the annihilator\(^9\) of \( L \). In the language of projective geometry the mapping \( L \mapsto L^\perp \) is a correlation. It is an involution in the following sense. For any subspace \( N \subset E^* \) we define its annihilator in \( E \) to be the subspace

\[
N_\perp = \{ x: x \in E, f(x) = 0 \ (f \in N) \}.
\]

Then\(^{10} \) \((L^\perp)_\perp = L \). We note the further obvious implications

\[
L_1 \subset L_2 \Rightarrow L^\perp_1 \supset L^\perp_2, \quad N_1 \subset N_2 \Rightarrow (N_1)_\perp \supset (N_2)_\perp
\]

and the correlation relations

\[
\left( \sum_{\sigma} L^{\perp}_{\sigma} \right) = \bigcap_{\sigma} L^{\perp}_{\sigma}, \quad \left( \sum_{t} N_t \right)_\perp = \bigcap_t (N_t)_\perp
\]

for any families of subspaces \( \{L_\sigma\} \) and \( \{N_t\} \) in \( E \) and \( E^* \) respectively.

With each homomorphism \( h: E_1 \to E_2 \) are associated two subspaces: its kernel \( \text{Ker} \ h \subset E_1 \), the set of solutions of the homogeneous equation \( hx = 0 \), and its image \( \text{Im} \ h \), the set of those \( y \) for which the inhomogeneous equation \( hx = y \) is solvable. The image \( \text{Im} \ h \) is isomorphic to the factor space \( E_1/\text{Ker} \ h \) (theory of E. Noether); the canonical isomorphism is \( h_{[x]} = hx \).

**Example.** Let us consider the homomorphism \( r: E^* \to L^* \) \((L \) is a subspace of \( E \)) obtained by restricting functionals to \( L \): \( rf = f|_L \). Clearly \( \text{Ker} \ r = L^\perp \),

---

\(^8\)In this case we have the formula

\[
\dim E^* = (\text{card} \ K) \text{dim} E.
\]

\(^9\)This definition remains meaningful for any non-empty subset \( X \subset E \) and as before \( X^\perp \) is a subspace of \( E^* \).

\(^{10}\)In general \((N_\perp)_\perp \neq N \) although \( N \subset (N_\perp)_\perp \).
Im $r = L^*$. Consequently $L^* \simeq E^*/L^\perp$. It is appropriate to point out here that $(E/L)^* \simeq L^\perp$ (the canonical isomorphism is $\varphi \mapsto \varphi j$, where $\varphi \in (E/L)^*$ and $j : E \rightarrow E/L$ is the natural homomorphism).

Returning to our consideration of kernels and images, we note that injectivity (monomorphicity) of a homomorphism $h$ is equivalent to the identity $\ker h = 0$ and surjectivity (epimorphicity) is equivalent to the identity $E_2/\text{Im } h = 0$. This factor space is called the cokernel and is denoted by $\text{Coker } h$ (and $E_1/\ker h$ is the coimage, denoted by $\text{Coim } h$).

The dimensions of the subspaces $\ker h, \text{Im } h$ are called respectively the defect and rank and are denoted by $\text{def } h, \text{rk } h$; their codimensions are the codefect (codef $h$) and corank (cork $h$). Clearly, as a consequence of Noether's theorem, we have $\text{codef } h = \text{rk } h$.

Now let us consider the extremely important canonical homomorphism $v : E \rightarrow E^{**}$ which is defined by the formula

$$(v)(f)(x) = f(x) \quad (x \in E, f \in E^*).$$

It is injective since, if $f(x) = 0$ for all $f \in E^*$, then $x = 0$. Identifying $\text{Im } v$ with $E$ we obtain the inclusion $E \subset E^{**}$. The space $E$ is said to be algebraically reflexive if $E^{**} = E$.

**Theorem.** In order that $E$ be algebraically reflexive it is necessary and sufficient that it be finite-dimensional.

**Proof.** If $E$ is infinite-dimensional then $\dim E^{**} > \dim E$. □

We note that as well as its small annihilator $N^\perp \subset E$ any subspace $N \subset E^*$ also has a big annihilator: $N^\perp \subset E^{**}$. Moreover $N^\perp = N^\perp \cap E$.

**Example 1.** If $N = 0$ then $N^\perp = E^*$ and $N^\perp = E^{**}$.

**Example 2.** If $N = E^*$ then $N^\perp = 0$ and $N^\perp = 0$.

With each homomorphism $h : E_1 \rightarrow E_2$ there is associated its algebraic conjugate $h^* : F_2^* \rightarrow F_1^*$ which is defined by the formula

$$(h^*f)(x) = f(hx) \quad (x \in E_1, f \in E_2^*).$$

There is a certain symmetry ('duality') between $h$ and $h^*$ which is reflected, in particular, in the following system of formulae:

$$\ker h^* = (\text{Im } h)^\perp, \quad \ker h = (\ker h^*)^\perp; \quad (3)$$

$$\text{Im } h^* = (\text{Ker } h)^\perp, \quad \text{Im } h = (\text{Im } h^*)^\perp. \quad (4)$$

The result that orthogonality of the vector $y$ to all solutions of the adjoint homogeneous equation $h^*f = 0$ is a criterion for solvability of the equation $hx = y$ is contained in these. We see also that injectivity (surjectivity) of a homomorphism $h$ is equivalent to surjectivity (injectivity) of the conjugate homomorphism $h^*$.

In addition to formulae (3) and (4) we note the relations:

$$\ker h^* \simeq (\text{Coker } h)^*, \quad \text{Coker } h^* \simeq (\ker h)^*, \quad \text{Im } h^* \simeq (\text{Im } h)^*.$$
Substitution of \( h^\# \) for \( h \) in the left members of formulae (3) and (4) leads to the identities

\[
\text{Im } h^{\#\#} = [(\text{Im } h)^{\perp}]^{\perp}, \quad \text{Ker } h^{\#\#} = [(\text{Ker } h)^{\perp}]^{\perp}.
\]

The homomorphism \( h^{\#\#} : E_1^{\#\#} \to E_2^{\#\#} \) is the natural extension of the homomorphism \( h \) from \( E_1 \) to \( E_1^{\#}\).

The linear spaces over a given field \( K \) are the objects of a category in which the homomorphisms serve as morphisms. The following contravariant adjoint functor\(^{11}\) acts on this category: \( E \mapsto E^*, \, h \mapsto h^\# \, ((gh)^* = h^\# g^\#, \, 1^* = 1) \). The more general contravariant functor \( E \mapsto \text{Hom}(E, E_0) \) (\( E_0 \) is a fixed space) transforms the morphisms by means of the commutative diagram

\[
\begin{array}{ccc}
E_1 & \xrightarrow{g} & E_2 \\
\downarrow & & \downarrow \\
E_0 & & \\
\end{array}
\]

The covariant functor \( E \mapsto \text{Hom}(E_0, E) \) is constructed similarly.

With each linear space \( E \) there is associated the set \( \text{End } E \equiv \text{Hom}(E, E) \equiv L(E) \) of endomorphisms or linear operators on \( E \) (defined on the whole of \( E \)). Under all the available algebraic structures on \( L(E) \) it is an associative (non-commutative if \( \dim E > 1 \)) algebra with identity \( 1 \) over the field \( K \). The invertible elements of this algebra, i.e. the automorphisms of the space \( E \), form a group \( \text{Aut } E \equiv GL(n, K) \).

The idempotents of the algebra \( L(E) \), i.e. operators \( P \) satisfying the condition \( P^2 = P \), are closely connected with the geometry of subspaces; in fact in this case

\[
F = \text{Im } P + \text{Ker } P, \quad P|_{\text{Im } P} = 1
\]

and conversely if \( E = L + M \) the representation \( x = u + v \, (u \in L, \, v \in M) \) of an arbitrary vector \( x \in E \) defines a linear operator \( P (Px = u) \) for which \( P^2 = P \), \( \text{Im } P = L \) and \( \text{Ker } P = M \). This is the projection onto \( L \) along \( M \) and \( 1 - P \) is the projection onto \( M \) along \( L \). If \( \{u_i\} \) is a basis in \( L \), \( \{v_j\} \) is a basis in \( M \) and \( \xi_i \) is the coordinate of the vector \( x \) with respect to the basis \( \{u_i\} \cup \{v_j\} \) in \( E \) corresponding to the vector \( u_i \), then

\[
Px = \sum_i \xi_i u_i.
\]

Associated with any decomposition of a space \( E \) into the direct sum of subspaces \( L_\alpha \) is a family of projections \( P_\alpha \) which are defined by the conditions

\[
\text{Im } P_\alpha = L_\alpha, \quad \text{Ker } P_\alpha = \sum_{\tau \neq \alpha} L_\tau.
\]

Clearly for any \( x \in E \) there is only a finite set of its projections \( P_\alpha x \) different from

\(^{11}\) In addition the mapping \( h \mapsto h^\# \) is linear:

\[
(h_1 + h_2)^* = h_1^* + h_2^*, \quad (zh)^* = zh^*.
\]
We note that the projections under consideration are pairwise annihilating, i.e. \( P_{\tau} P_{\sigma} = 0 \) \((\tau \neq \sigma)\) and the family \( P_{\sigma} \) is total in the sense that
\[
\bigcap_{\sigma} \text{Ker} P_{\sigma} = 0.
\]

Invariant subspaces of any linear operator \( A: E \to E \) can be conveniently characterised in terms of projections. In order that a subspace \( L \) be invariant \(^{12}\) (i.e. \( AL \subset L \)) it is necessary that any projection \( P \) onto \( L \) should satisfy the relation \( PAP = AP \) and it is sufficient that even just one projection should satisfy this relation. The stronger commutativity condition \( PA = AP \) implies the invariance of not only the subspace \( L = \text{Im} P \) but also of its complement \( M = \text{Ker} P \). In this case we say that the subspace \( L \) is completely invariant under (or reduces completely) the operator \( A \), about which we say in turn that it decomposes into the direct sum of the restrictions \( A|_{L} \) and \( A|_{M} \). This construction extends in an obvious way to the situation where the space \( E \) is decomposed into a direct sum of an arbitrary family of invariant subspaces.

The famous problem of the existence of non-trivial (i.e. different from 0, \( E \)) invariant subspaces is trivial in the purely algebraic context.

**Theorem.** If the field \( K \) is algebraically closed and \( \dim E > 1 \) then any linear operator \( A: E \to E \) has a non-trivial invariant subspace.

**Proof.** We can assume that \( E \) is infinite-dimensional and that \( A \) does not have any finite-dimensional invariant subspaces other than the zero subspace. Then for any \( x \in E, \; x \neq 0 \) the system \( \{ A^{k}x \}_{k=0}^{\infty} \) is linearly independent. The desired subspace is \( \text{Lin} \{ A^{k}x \}_{k=0}^{\infty} \).

In certain cases it is easy to describe all invariant subspaces.

**Example.** Let us consider differentiation as an operator \( D \) on the space \( \Pi \) of all polynomials in one variable. The invariant subspaces are precisely the subspaces \( \Pi_{n} \) of polynomials of degree at most \( n \) \((n = 0, 1, 2, \ldots)\), the zero subspace \( 0 \) and \( \Pi \). The operator \( D \) is unicellular in the sense that its invariant subspaces form a chain under inclusion: \( 0 \subset \Pi_{0} \subset \Pi_{1} \subset \cdots \subset \Pi \). Consequently \( D \) has no non-trivial completely invariant subspaces.

If the subspace \( L \) is invariant for the linear operator \( A \), then it is invariant for all polynomials \( P(A) \). This obvious fact has a partial converse (Yu.I. Lyubich, 1955).

**Theorem.** Let the subspace \( L \subset E \) be invariant for the operator \( Q(A) \) where \( Q \) is some polynomial of degree \( n \). Let us suppose that there exists a set \( \{ \lambda_{1}, \ldots, \lambda_{n} \} \subset K \) such that

\(^{12}\) We also say that \( L \) reduces the operator \( A \).
1) the equations $A\mathbf{u} - \lambda_k \mathbf{u} = \mathbf{v} (1 \leq k \leq n)$ are solvable in $L$ for all $\mathbf{v} \in L$,
2) the equation $\Delta(A)d = 0$, where $\Delta(\lambda) = \prod_{1 \leq k \leq n} (\lambda - \lambda_k)$, has only the trivial solution in $L$.

Then $L$ is invariant for $A$.

In fact it is shown under the stated conditions that if $x \in L$ and $Q(A)x \in L$ then $A^k x \in L (1 \leq m \leq n)$. It is even possible to relax the requirement that $A$ is defined on the whole of $E$, but then it is necessary to have $x \in D(A^n)$. The following well-known result in analysis is a consequence of this general formulation.

**Lemma of Esclangon.** Let $x(t) (t \geq 0)$ be a bounded solution of an $n$th order linear differential equation with constant coefficients and bounded right hand side. Then all the derivatives $x^{(m)}(t) (1 \leq m \leq n)$ are also bounded.

A further development of the approach described leads to bounds for the intermediate derivatives which are similar to A.N. Kolmogorov's inequalities.

### 1.3. The Algebraic Theory of the Index.

A homomorphism $h: E_1 \to E_2$ is said to be *Noetherian* if $\text{Ker} h$ and $\text{Coker} h$ are finite-dimensional. If $h$ is a Noetherian homomorphism its *index* is defined as

$$\text{ind} h = \text{def} h - \text{cork} h = \dim(\text{Ker} h) - \dim(\text{Coker} h).$$

A Noetherian homomorphism with zero index is called a *Fredholm homomorphism*. In this case surjectivity is equivalent to injectivity, i.e. the *Fredholm alternative* holds: either the inhomogeneous equation $hx = y$ is solvable for all $y$ or the homogeneous equation $hx = 0$ has a non-trivial solution.

It follows from formulae (3) and (4) that the homomorphism $h^*$ is Noetherian if and only if the homomorphism $h$ is Noetherian and in the Noetherian case

$$\text{def} h^* = \text{cork} h, \quad \text{cork} h^* = \text{def} h,$$

so that $\text{ind} h^* = - \text{ind} h$. It is appropriate to note here that

$$\text{ind} h = \text{def} h - \text{def} h^*.$$

The product $gh$ of two Noetherian homomorphisms is Noetherian and its index has the *logarithmic property*:

$$\text{ind}(gh) = \text{ind} g + \text{ind} h.$$

Consequently the product of two Fredholm homomorphisms is also a Fredholm homomorphism. This property is preserved under 'small' perturbations. The following definition is required for a more precise formulation: a homomorphism is said to be of *finite rank* if its image is finite-dimensional.

**Theorem.** If $A: E \to E$ is an operator of finite rank then $1 - A$ is a Fredholm operator.

**Proof.** We take a complement of the subspace $L = \text{Im} A$ in the whole space: $L \oplus M = E$. If $x = u_0 + v (u_0 \in L, v \in M)$ then $x = u + (1 - A)v$, where $u = \ldots$
Chapter 2. Foundations and Methods

Consequently \( L + (1 - A)M = E \) and it is easy to see that this decomposition is direct. Let us consider the operator \( \tilde{A} = A|_L \) (clearly the subspace \( L \) is invariant under \( A \)). From what has gone before, \( \text{Coker}(1 - A) \cong \text{Coker}(1 - \tilde{A}) \). At the same time \( \text{Ker}(1 - A) = \text{Ker}(1 - \tilde{A}) \). Consequently \( 1 - A \) is Noetherian and \( \text{ind}(1 - A) = \text{ind}(1 - \tilde{A}) = 0 \), since every operator on a finite-dimensional space is a Fredholm operator.

The operators of finite rank on \( E \) form a two-sided ideal \( \Phi \) in the algebra \( L(E) \). Thus if \( U \in \text{Aut} E \) and \( A \in \Phi \) then \( U - A = U(1 - U^{-1}A) \) is a Fredholm operator.

It is useful to record the general form of an operator of finite rank. It is

\[
A = \sum_{k=1}^{m} f_k(\cdot) u_k,
\]

where \((u_k)\) is an arbitrary system of vectors and \((f_k)\) is an arbitrary system of linear functionals. For the \((u_k)\) we can always choose a basis in \( \text{Im} A \) and then the system \((f_k)\) is linearly independent. If these systems are biorthogonal (and only in this case) \( A \) is a projection.

In a certain sense the investigation of arbitrary Noetherian operators reduces to the previous situation.

**Theorem.** If the operator \( T: E \to E \) is Noetherian there exist operators \( R_l \) and \( R_r \) (left and right regularisers) such that

\[
R_l T = 1 - A, \quad TR_r = 1 - B,
\]

where \( A, B \) have finite rank. Conversely, any operator which allows both of the stated regularisations is Noetherian.

**Proof.** If \( A \) is a projection onto \( \text{Ker} T \) and \( P \) is a projection onto \( \text{Im} T \) and if \( \tilde{T} = T|_{\text{Ker} A} \) (clearly \( \tilde{T}: \text{Ker} A \to \text{Im} T \) is an isomorphism), then \( R_l = (1 - A)\tilde{T}^{-1}P \), \( R_r = \tilde{T}^{-1}P \) \((B = 1 - P) \). The converse statement follows from the previous theorem and the inclusions \( \text{Ker} T \subset \text{Ker}(1 - A), \text{Im} T \supset \text{Im}(1 - B) \).

**Remark.** It can be said that the Noetherian operators on \( E \) are invertible modulo \( \Phi \) (moreover \( R_l \equiv R_r \) (mod \( \Phi \)).

It follows from the logarithmic property of the index that

\[
\text{ind} T = -\text{ind} R_l = -\text{ind} R_r.
\]

Suppose that the operator \( V: E \to E \) is small in the sense that \( 1 + R_l V \) and \( 1 + VR_r \) are invertible. Then

\[
(T + V)R_r = (1 - B(1 + VR_r)^{-1})(1 + VR_r),
\]

and

\[
R_l(T + V) = (1 + R_l V)(1 - (1 + R_l V)^{-1}A)
\]

are Fredholm operators, from which it follows that \( T + V \) is Noetherian and

\[
\text{ind}(T + V) = -\text{ind} R_l = \text{ind} T.
\]
From this emerges the important fact of the stability of the index under small perturbations which is entirely natural since the index takes integral values.

The calculation of the index of some particular operators is usually an extremely interesting and often very difficult problem, as for example in the case of the famous Atiyah-Singer theorem on the index of an elliptic linear differential operator.

1.4. Systems of Linear Equations. Let \( F = (f_i)_{i \in I} \) be an arbitrary system of linear functionals on the space \( E \) (the index set \( I \) can be finite or infinite). Let us consider the system of linear equations

\[
 f_i(x) = \eta_i \quad (i \in I),
\]

where the \( \eta_i \) are given scalars. Clearly a necessary condition for solvability (consistency) of the system (5) is that the implication

\[
 \sum_i \lambda_i f_i = 0 \Rightarrow \sum_i \lambda_i \eta_i = 0
\]

holds for any system of scalars \( (\lambda_i)_{i \in I} \) with finite support. Conversely, if (6) is satisfied the system (5) is equivalent to a subsystem corresponding to a maximal linearly independent subset \( F_0 \subset F \).

If we introduce the homomorphism \( h_F: E \to \Phi(I, K) \) by means of the formulae

\[
 (h_Fx)(i) = f_i(x) \quad (i \in I)
\]

we can write the system (5) in the form of an equation

\[
 h_Fx = y,
\]

where \( y(i) = \eta_i \) (\( i \in I \)).

**Lemma.** If \( f_1, \ldots, f_m \) are linearly independent linear functionals, the system of equations \( f_i(x) = \eta_i \) (\( 1 \leq i \leq m \)) is solvable for any right hand sides \( \eta_1, \ldots, \eta_m \).

**Proof.** Because the set \( M = \{1, \ldots, m\} \) is finite we can identify \( \Phi(M, K) \# \) with \( \Phi(M, K) \). The conjugate homomorphism \( h_F^* \) maps any collection \( (\lambda_i) \) to \( \sum_{i=1}^m \lambda_i f_i \). By the condition \( h_F^* \) is injective. Consequently \( h_F \) is surjective.

**Corollary.** If the system (5) has finite rank (i.e. \( \text{rk} F \) is finite) then for its consistency it is necessary and sufficient that condition (6) be satisfied.

In the case of infinite rank condition (6) is no longer sufficient.

**Example.** Let \( \{e_i\}_{i \in I} \) be a basis in \( E \) and \( \{\xi_i\}_{i \in I} \) the set of coordinate functionals. Because of their linear independence the system of equations \( \xi_i(x) = \eta_i (i \in I) \) satisfies condition (6) but it is solvable only for those \( (\eta_i)_{i \in I} \) of finite support.

In the general case, (6) is necessary and sufficient for the consistency of each finite subsystem of system (5), i.e. for the reduction to the case of linearly independent equations.

If the system (5) is consistent and \( x_0 \) is some particular solution of it, then the general solution is described by the formula \( x = x_0 + z \), where \( z \) is the general solution of the corresponding homogeneous system:

\[
 f_i(z) = 0 \quad (i \in I).
\]

(50)
We denote the subspace of solutions of system (5) by $N_F$. If $\{z_j\}_{j \in J}$ is a basis of it, the general solution of the system (5) has the form

$$x = x_0 + \sum_j \alpha_j z_j,$$

where $(x_j)_{j \in J}$ runs through all possible scalar systems with finite support.

For uniqueness of the solution of system (5) it is necessary and sufficient that $N_F = 0$, i.e. system (5) should have only the trivial solution. In this case the system of functionals $F$ is said to be total.

Since $N_F = \text{Ker } h_F$ we have

$$\text{codim } N_F = \text{rk } h_F.$$

In the case where the system consists of one equation $f(x) = 0$ ($f \neq 0$) we have $\text{codim } N_F = 1$, i.e. $N_F$ is a hyperplane ($N_F = \text{Ker } f$). Conversely, for any hyperplane $H \subset E$ there exists a linear functional $f \neq 0$ such that $\text{Ker } f = H$. In fact, since $\text{dim}(E/H) = 1$, we can identify $E/H$ with the field $K$ and then $f: E \rightarrow E/H$ is the natural homomorphism. Finally, if $\text{Ker } f_1 = \text{Ker } f_2$ then $f_1$ and $f_2$ are proportional. Thus the set of hyperplanes in the space $E$ can be identified with the projective space $\mathbb{P}E^*$ consisting of the classes under proportionality on $E^* \setminus \{0\}$.

In contrast to the very concise representation of the system (5) in the form of the equation $h_F x = y$ we can represent it in scalar form

$$\sum_{j \in J} \alpha_{ij} \xi_j = \eta_i \quad (i \in I),$$

(7)

where $(\xi_j)_{j \in J}$ are the coordinates of the vector $x \in E$ with respect to some basis $\{e_j\}_{j \in J}$ and $\alpha_{ij} = f_i(e_j)$. In addition only solutions of finite support are sought. However we can dispense with this restriction if the rows in (7) have finite support. From the geometric point of view a system of the form

$$f(x_i) = \eta_i \quad (i \in I)$$

(5*)

is obtained, where $x_i$ is the vector with coordinates $(\alpha_{ij})_{j \in J}$ and $f$ is an unknown linear functional. The condition

$$\sum_i \lambda_i x_i - 0 \Rightarrow \sum_i \lambda_i \eta_i = 0 \quad (\lambda \in \Phi_o(I, K))$$

(6*)

is sufficient for solvability of the system (5*) since, if the system $(x_i)_{i \in I}$ is already linearly independent, it can be extended to a basis, after which the functional $f$ is defined by the values $\eta_i$ on the vectors $x_i$ and is extended arbitrarily on the other basis vectors. With this we have established

**Theorem of Toeplitz.** For consistency of the system of equations (7) when in each equation the set of non-zero coefficients is finite\(^{13}\), it is necessary (trivial) and sufficient that all its finite subsystems be consistent.

\(^{13}\)We shall sometimes call such a system a Toeplitz system.
It is interesting that, under specific conditions on the field $K$, this theorem generalises\(^{14}\) to systems of algebraic (non-linear) equations (see Section 1.7).

1.5. Algebraic Operators. A linear operator $A: E \to E$ is said to be *algebraic* if there exists an annihilating polynomial $\varphi \neq 0$ for $A$: $\varphi(A) = 0$. In this case an annihilating polynomial of least degree is uniquely determined up to a scalar factor. We can therefore take it to be monic and then it is called the *minimal polynomial* of the operator $A$. All annihilating polynomials are divisible by the minimal polynomial.

**Example 1.** Any projection $P$ is algebraic: $P^2 = P$. If $P \neq 0$, 1, its minimal polynomial is $\lambda^2 - \lambda$.

**Example 2.** If $A$ is a nilpotent operator, i.e. $A^m = 0$ for some $m$, then it is clearly algebraic and its minimal polynomial has the form $\lambda^r$ where $r \leq m$. The number $r$ is called the *index of nilpotence* of the operator $A$.

If $E$ is finite-dimensional every linear operator on $E$ is algebraic. This property is fundamental to the classical theorem of Jordan, thus allowing the theorem to be generalised to algebraic operators on an infinite-dimensional space (whose minimal polynomials factorise into linear factors over the field $K$).

The set of eigenvalues of an algebraic operator $A$ is called its *spectrum* and is denoted by $\text{spec } A$. Associated with each $\lambda \in \text{spec } A$ are the *root subspaces*

$$W_k(\lambda) = \{x: (A - \lambda 1)^k x = 0\}$$

of orders $k = 1, 2, 3, \ldots$. They are all invariant under $A$. The subspace $W_1(\lambda)$ is the eigenspace and clearly

$$0 \subsetneq W_1(\lambda) \subset W_2(\lambda) \subset \cdots .$$

The union

$$W(\lambda) = \bigcup_{k=1}^{\infty} W_k(\lambda)$$

is called the *maximal root subspace at the point* $\lambda$ (correspondingly, the vectors $x \in W(\lambda) \setminus \{0\}$ are called *root vectors*). If there exists $r \geq 1$ such that $W(\lambda) = W_r(\lambda)$ then the least such $r$ is called the *order* (ord $\lambda$) of the eigenvalue $\lambda$; otherwise ord $\lambda = \infty$. It is easy to see that, if ord $\lambda = r < \infty$, then in the chain of root subspaces the first $r - 1$ inclusions are strict and the rest are equalities.

The operator $A|_{W_r(\lambda)}$ has a *Jordan structure*, i.e. it decomposes into a direct sum (in general of an infinite set\(^{15}\)) of finite-dimensional invariant subspaces on each of which the matrix of the operator with respect to a suitable basis is a *Jordan cell*

\(^{14}\)This range of questions is closely connected with the well-known G"{o}del-Mal'tsev theorem on compactness in mathematical logic.

\(^{15}\)For finiteness of this set it is necessary and sufficient that the eigenspace $W_1(\lambda)$ be finite-dimensional (then all the $W_k(\lambda)$ are finite-dimensional).
of order \( \leq r \). Jordan's theorem is a combination of this fact with the following important proposition.

**The Spectral Theorem.** Let \( A \) be an algebraic operator and let its minimal polynomial have the form

\[
\varphi(\lambda) = \prod_{j=1}^{s} (\lambda - \lambda_j)^{r_j},
\]

where \( \lambda_1, \ldots, \lambda_s \) are pairwise distinct. Then \( \{\lambda_j\} \) is the set of all eigenvalues of the operator \( A \), their orders are equal to the corresponding \( r_j \) and the space \( E \) decomposes into the direct sum of the maximal root subspaces:

\[
E = W(\lambda_1) \oplus \cdots \oplus W(\lambda_s). \tag{8}
\]

In particular, if all the roots of the minimal polynomial are simple (and only in this case), the operator \( A \) is diagonalisable, i.e. \( E \) decomposes into the direct sum of the eigenspaces.

A decomposition of the identity operator \( \mathbf{1} \) into the sum of root projections \( P_1, \ldots, P_s \) is defined by the decomposition (8). These projections have the following explicit representation: \( P_k = \varphi_k(A) \), where \( \varphi_k \) is the interpolating polynomial corresponding to the conditions:

\[
\varphi_k(\lambda_j) = \delta_{jk}; \quad \varphi_k^{(i)}(\lambda_j) = 0 \quad (1 \leq i \leq r_j - 1; 1 \leq j \leq s).
\]

The root projections commute with the operator \( A \) and in the diagonalisable case

\[
A = \sum_{j=1}^{s} \lambda_j P_j
\]

(spectral resolution of the operator).

A further consequence of the spectral theorem is the invertibility of the operator \( A - \lambda \mathbf{1} \) for all \( \lambda \notin \text{spec } A \). The inverse operator \( (A - \lambda \mathbf{1})^{-1} \) is called the resolvent of the operator \( A \) and is denoted by \( R_\lambda(A) \) (abbreviated, \( R_\lambda \)). Correspondingly \( K \setminus \text{spec } A \) is called the resolvent set and its points are said to be regular for \( A \); it is therefore denoted by \( \text{reg } A \). We have a decomposition into partial fractions:

\[
R_\lambda(A) = \sum_{j=1}^{s} \sum_{i=0}^{r_j-1} \frac{P_j^{(i)}}{(\lambda - \lambda_j)^{i+1}},
\]
where $P_j^{(0)} = P_j$ are the root projections, and $P_j^{(i)} = (A - \lambda_j 1)^i P_j$. Consequently

$$R_\lambda(A) = \frac{\theta(\lambda, A)}{\phi(\lambda)},$$

where $\theta$ is a polynomial in two variables (deg $\theta \leq$ deg $\phi - 1$).

The operator $p(A)$ is naturally defined for any rational function $p$ whose poles are different from $\lambda_1, \ldots, \lambda_s$. Moreover the Lagrange-Sylvester formula holds:

$$\rho(A) = \sum_{j=1}^{s} \frac{\rho^{(0)}(\lambda_j)}{i!} P_j^{(i)}.$$

The correspondence $p \mapsto \rho(A)$ is called the functional calculus for the operator $A$. This is a homomorphism of the algebra of rational functions which are regular on $\text{spec } A$ into the algebra of operators. We note in this context the spectral mapping theorem: $\text{spec } \rho(A) = \rho(\text{spec } A)$.

Now let $A$ be any operator, $\{\lambda_j\}_{j=1}^s$ some collection of its eigenvalues, $\eta_j = \text{ord } \lambda_j$ ($1 \leq j \leq s$) and

$$q(A) = \prod_{j=1}^{s} (A - \lambda_j)^{\eta_j},$$

where $\eta_j < r_j + 1$ ($1 \leq j \leq s$). Let us consider the linear equation $\varphi(A)x = 0$ with unknown $x \in E$. The subspace of its solutions $N = \text{Ker } \varphi(A)$ is invariant under $A$ and the operator $A|_N$ is algebraic. Its minimal polynomial coincides with $\varphi$. By the spectral theorem

$$N = W_{\eta_1}(\lambda_1) + \cdots + W_{\eta_s}(\lambda_s)$$

(thes subspace, which are root subspaces for $A$, are contained in $N$) and, in more detail, $A|_N$ has a Jordan structure.

We will discuss two classical examples from this point of view.

**Example 1.** The differentiation operator $D = d/dt$ on the complex function space $C^\infty(\mathbb{R})$ has all $\lambda \in \mathbb{C}$ for its eigenvalues (of infinite order). The root subspace $W_{\eta}(\lambda)$ consists of all functions of the form $\pi(t) e^{\lambda t}$, where $\pi$ is a polynomial of degree $\leq k - 1$.

If follows from the spectral theorem that the general solution of the linear differential equation with constant coefficients

$$\varphi(D)x = 0 \quad \left( \varphi(\lambda) = \prod_{j=1}^{s} (\lambda - \lambda_j^{\eta_j}) \right)$$

has the form

$$x(t) = \sum_{j=1}^{s} \pi_j(t) e^{\lambda_j t},$$

where the $\pi_j$ are arbitrary polynomials of degrees $\leq \eta_j - 1$ ($1 \leq j \leq s$).

**Example 2.** Let us consider the shift operator

$$(Tx)_n = \xi_{n+1} \quad (n = 0, 1, 2, \ldots)$$
on the space $s$ of all complex sequences $x = (\xi_n)_{n=0}^\infty$. As in the previous example, its eigenvalues fill out the whole plane $C$ and have infinite order. For $\lambda \neq 0$ the root subspace $W_k(\lambda)$ consists of all sequences of the form $\{\pi(n)\xi_n\}_{n=0}^\infty$, where $\pi$ is a polynomial of degree $\leq k - 1$. Consequently the general solution of the linear difference equation with constant coefficients

$$\varphi(T)x = 0 \quad \left(\varphi(\lambda) = \prod_{j=1}^s (\lambda - \lambda_j)^{\nu_j}; \quad \varphi(0) \neq 0\right)$$

has the form

$$\xi_n = \sum_{j=1}^s \pi_j(n)\lambda_j^n \quad (n = 0, 1, 2, \ldots),$$

where the $\pi_j$ are arbitrary polynomials of degrees $\leq \nu_j - 1 \ (1 \leq j \leq s)$.

### 1.6. General Principles of Summation of Series

In this Section we will indicate a further interesting application of the methods of linear algebra to problems of analysis. Let $L$ be any subspace of the space $s$ of all complex sequences $x = (\xi_n)_{n=0}^\infty$ which is invariant under the shift operator $T$. The following definition (or if preferred, axiom of summation) is due to A.N. Kolmogorov (1925).

A linear functional $\sigma$ on $L$ is called a generalised sum (or a method of summation) of the series $\xi_0 + \xi_1 + \xi_2 + \cdots$ if it satisfies the equation

$$\sigma(Tx) - \sigma(x) = -\xi_0 \quad (x = (\xi_n)_{n=0}^\infty \in L)$$

(i.e. the equation $(T|_L)^\#\sigma - \sigma = -\xi_0$, where $\xi_0$ is regarded as a linear functional on $L$). Further, we shall say that $L$ admits summation if a generalised sum exists on it. We denote the class of such subspaces by $\Sigma$.

On the subspace $C$ of those $x$ for which the series $\xi_0 + \xi_1 + \xi_2 + \cdots$ converges the ordinary sum is clearly a generalised sum. Many classical methods of summation on their natural domains of definition (in particular, the Cesàro method) satisfy condition (9).

A method of summation on $L \supset C$ is said to be regular if it coincides on $C$ with the ordinary sum. Such a method always exists if $L \in \Sigma$.

We denote by $N$ the subspace of sequences of the form $(\pi(n))_{n=0}^\infty$ where $\pi$ runs through all polynomials. This is the maximal root subspace of the operator $T$ for $\lambda = 1$.

**Theorem 1.** Let $L$ be an invariant subspace. Then $L \in \Sigma$ if and only if $L \cap N = 0$.

We call the subspace $M$ $\Sigma$-maximal if $M \in \Sigma$ and there is no extension $\tilde{M} \supset M$ in the class $\Sigma$ which preserves the generalised sum on $M$. Every $L \in \Sigma$ can be extended to a $\Sigma$-maximal subspace in such a way that the generalised sum on $L$ is preserved.

**Theorem 2.** Let $M$ be an invariant subspace. Then $M$ is $\Sigma$-maximal if and only if it is a complement of the subspace $N$ in the whole space $s$.

Naturally, Zorn's lemma is used in the proofs of the results listed.
If \( L \in \Sigma \) uniqueness of the generalised sum on \( L \) is equivalent to the absence on \( L \) of linear functionals \( \rho \neq 0 \) which are invariant with respect to shift: 
\[
\rho(Tx) = \rho(x) \quad (x \in L).
\]
This in turn is equivalent to the surjectivity of the operator \((T - 1)|_L\).

The generalised sum is not unique even on the subspace \( l^1 \), i.e. even under the condition of absolute convergence: 
\[
|\xi_0| + |\xi_1| + \cdots < \infty.
\]
However it can be shown that on \( l^1 \) there is a unique non-negative generalised sum, i.e. \( \sigma(x) \geq 0 \) if \( x \geq 0 \) (i.e. if \( \xi_n \geq 0 \) \((n = 0, 1, 2, \ldots))\); it must therefore coincide with the ordinary sum.

For any \( L \) we put \( L_+ = \{x: x \in L, x \geq 0\} \).

**Theorem 3.** A non-negative generalised sum exists on a subspace \( L \in \Sigma \) if and only if \( L_+ \subset l^1 \).

Let us now consider the invariant subspace \( Q \) of all sequences of the form 
\[
\xi_n = \sum_{j=1}^{n} \pi_j(n)\lambda_j^n,
\]
where all the \( \lambda_j \) are different from 1 and 0 and the \( \pi_j \) are polynomials. Since \( Q \cap N = 0 \), we have \( Q \in \Sigma \). The operator \( T - 1 \) on \( Q \) is surjective and so the generalised sum on \( Q \) is unique. It can be described in a straightforward manner. To this end we note that if \( x \in Q \) the power series 
\[
\sum_{n=0}^{\infty} \xi_n z^n
\]
converges on some disk \( |z| < r_x \). Its ordinary sum \( f_x(z) \) (the generating function of the sequence \( x \)) is rational and its poles are \( \lambda_j^{-1} \) \((1 \leq j \leq m)\).

**Theorem 4.** The unique generalised sum on \( Q \) is defined by the formula 
\[
\sigma(x) = f_x(1).
\]

The non-negativity requirement is not satisfied here. For example, if \( x = (\lambda^n)_{\infty} \) where \( \lambda > 0 \) \((\lambda \neq 1)\), then \( \sigma(x) = (1 - \lambda)^{-1} \).

In conclusion we note that \( Q \) is contained in all \( \Sigma \)-maximal subspaces. It is interesting that the generalised sum on each \( \Sigma \)-maximal subspace is unique.

1.7. **Commutative Algebra.** The term 'algebra' in this section means a commutative associative algebra with identity element \( e \neq 0 \) over some field \( K \). The algebra \( \Phi(S, K) \) of scalar functions on an arbitrary set \( S \), or any of its subalgebras, can serve as an example. The algebras \( \Phi(S, K) \) and all their subalgebras are called function algebras.

We stress that according to the definition of a subalgebra the identity \( e \) must belong to it. In contrast to this, ideals do not contain the identity (with the exception of the ideal which coincides with the whole algebra).

The presence of the identity allows us to consider the field \( K \) as being embedded in the algebra. In this way the identity of the field \( K \) is identified with \( e \) (and in general \( \alpha \equiv ne \)).

An ideal \( M \) of an algebra \( \mathcal{H} \) is said to be maximal if \( M \neq \mathcal{H} \) and there is no ideal \( \mathcal{M} \supset M \) which is different from \( M \) and \( \mathcal{H} \). With the help of Zorn's lemma it can be shown that any ideal \( I \neq \mathcal{H} \) is contained in some maximal ideal (theorem of Krull). Therefore in \( \mathcal{H} \) there exists at least one maximal ideal. An algebra with a unique maximal ideal is said to be local. A very simple example is provided.
by $\mathfrak{A} = K + E$ with multiplication $(x + x)(\beta + y) = x\beta(x, \beta \in K; x, y \in E)$. The subspace $E \subset \mathfrak{A}$ is the unique maximal ideal.

In any algebra $\mathfrak{A}$ each element $a$ generates the ideal $(a) = \{y: y = ax (x \in \mathfrak{A})\}$. The set of non-invertible elements therefore coincides with the union of all maximal ideals. Consequently an algebra $\mathfrak{A}$ is a field if and only if its unique maximal ideal is the zero ideal. If $I \neq \mathfrak{A}$ is an ideal in $\mathfrak{A}$ then the factor algebra $\mathfrak{A}/I$ is a field if and only if the ideal $I$ is maximal.

The radical of an algebra $\mathfrak{A}$ is the set $\text{Rad } \mathfrak{A}$ of those elements $a \in \mathfrak{A}$ such that $e - ax$ is invertible for all $x \in \mathfrak{A}$.

**Theorem.** The radical of an algebra $\mathfrak{A}$ coincides with the intersection of all its maximal ideals.

**Proof.** If $M$ is a maximal ideal and $a \in M$ then $e - ax \notin M$ and if this holds for all $M$ it follows that $e - ax$ is invertible for all $x \in \mathfrak{A}$, i.e. $a \in \text{Rad } \mathfrak{A}$. Conversely let $a \in \text{Rad } \mathfrak{A}$ and suppose that $a \notin M$ for some maximal ideal $M$. Then the image of the element $a$ in the field $\mathfrak{A}/M$ is invertible, i.e. there exists $x \in \mathfrak{A}$ such that $ax \equiv e \pmod{M}$. But then $e - ax \in M$ which implies the non-invertibility of this element. \(\square\)

**Corollary 1.** $\text{Rad } \mathfrak{A}$ is an ideal.

An algebra $\mathfrak{A}$ is said to be **semisimple** if $\text{Rad } \mathfrak{A} = 0$.

**Corollary 2.** For any algebra $\mathfrak{A}$ the factor algebra $\mathfrak{A}/\text{Rad } \mathfrak{A}$ is semisimple.

If the element $a \in \mathfrak{A}$ is nilpotent, i.e. $a^m = 0$ for some $m$, then $e - a$ is invertible:

$$(e - a)^{-1} = \sum_{k=0}^{m-1} a^k.$$  

But if $a$ is nilpotent so also is $ax$ for any $x \in \mathfrak{A}$. Therefore all nilpotent elements are contained in the radical of the algebra. They form an ideal called the nil radical. The nil radical coincides with the intersection of all simple ideals (an ideal $I \neq \mathfrak{A}$ is said to be simple if there are no divisors of zero in the factor algebra $\mathfrak{A}/I$). If $E$ is finite-dimensional its nil radical coincides with its radical.

For each maximal ideal $M$ in an algebra $\mathfrak{A}$ the field $\mathfrak{A}/M$ is an extension of the field $K$, for the latter is canonically embedded in $\mathfrak{A}/M$ by means of the chain of natural homomorphisms$^{17}$:

$$K \rightarrow \mathfrak{A} \rightarrow \mathfrak{A}/M.$$  

We call an algebra spectral if $\mathfrak{A}/M = K$ for all $M$, i.e. the codimensions of all maximal ideals are equal to one. This condition is a fundamental axiom of the

---

$^{16}$ In general if $\mathfrak{A}_0$ is an algebra without identity the construction $\mathfrak{A} = K + \mathfrak{A}_0$ with multiplication $(x + x)(\beta + y) = x\beta + (xy + \beta x + xy)$ is called the unitisation of $\mathfrak{A}_0$.

$^{17}$ We stress that a homomorphism of algebras with identities must by definition map identity to identity.
general *spectral theory* set out below, the principal prototypes of which are to be found in the spectral theory of operators, in the theory of commutative Banach algebras and in algebraic geometry. In what follows the algebra $\mathfrak{A}$ is assumed to be spectral.

The *spectrum* of an element $a \in \mathfrak{A}$ is the set $\text{spec } a$ of those $\lambda \in K$ for which the element $a - \lambda e$ is not invertible. The remaining $\lambda \in K$ are said to be *regular*, the set of regular values is denoted by $\text{reg } a$ and it is called the *resolvent set* for $a$ since it is the domain of definition of the *resolvent* $R_{\lambda}(a) = (a - \lambda e)^{-1}$.

**Example.** If $a = ae$ then $\text{spec } a = \{a\}$, $R_{\lambda} = (a - \lambda e)^{-1}$. In particular $\text{spec } 0 = \{0\}$, $\text{spec } e = \{1\}$.

We note two fundamental relationships for the resolvent:

$$aR_{\lambda} = e + \lambda R_{\lambda}; \quad R_{\lambda} - R_{\mu} = (\lambda - \mu)R_{\lambda}R_{\mu}. $$

The second of these is called *Hilbert's identity* and is easily deduced from the first.

We denote by $\mathfrak{M}$ the set (space) of *maximal ideals* of an algebra $\mathfrak{A}$ and for each $M \in \mathfrak{M}$ we denote by $a(M)$ the image of the element $a \in \mathfrak{A}$ in $\mathfrak{A}/M$. Since $\mathfrak{A}/M = K$ we obtain a scalar function $\tilde{a}$ on $\mathfrak{M}$. For invertibility of the element $a$ it is necessary and sufficient that $\tilde{a}(M) \neq 0$ for all $M$, i.e. that $\tilde{a}$ be invertible in the function algebra $\Phi(\mathfrak{M}, K)$.

The mapping $a \mapsto \tilde{a}$ is an algebra homomorphism: $\mathfrak{A} \rightarrow \Phi(\mathfrak{M}, K)$. It is called the *canonical functional representation* of the algebra $\mathfrak{A}$. Its kernel coincides with $\text{Rad } \mathfrak{A}$ and its image $\mathfrak{A}$ is a subalgebra of $\Phi(\mathfrak{M}, K)$; moreover it is sufficiently rich in the sense that the functions in $\mathfrak{A}$ separate the points of the space $\mathfrak{M}$: if $M_1 \neq M_2$ then $M_1 \neq M_2$ and then for $a \in M_1 \setminus M_2$ we have $\tilde{a}(M_1) = 0$, $\tilde{a}(M_2) \neq 0$.

If $\mathfrak{A}$ is semisimple (and only in this case), $\mathfrak{A}$ is isomorphic to $K$. Thus any semisimple spectral algebra can be considered as an algebra of functions.

**Theorem.** *The set of values of the function $\tilde{a}$ coincides with the spectrum of the element $a$.***

**Proof.** We have:

$$\text{spec } a = \{\lambda : \exists M \in \mathfrak{M} \text{ s.t. } a - \lambda e \in M\} = \{\lambda : \exists M \in \mathfrak{M} \text{ s.t. } \tilde{a}(M) = \lambda\}. $$

**Corollary 1.** *The spectrum of any element is non-empty.*

Non-invertibility of an element $a$ implies that $0 \in \text{spec } a$. An element $a$ is said to be *quasinilpotent* if $\text{spec } a = \{0\}$.

**Corollary 2.** *The set of quasinilpotents coincides with the radical of the algebra.*

Thus any nilpotent is quasinilpotent. A more general result holds.

18 For this and the definitions which follow immediately it is not necessary that the algebra be spectral or even commutative.
Corollary 3. If \( a \) is an algebraic element and \( \varphi \) is its minimal polynomial\(^{19}\), then \( \varphi \) factorises over the field \( \mathbf{K} \) into linear factors and \( \text{spec} \ a \) coincides with the set of roots of the polynomial \( \varphi \).

Proof. It follows from \( \varphi(a) = 0 \) that \( \varphi(\psi(M)) = 0 \) and so \( \text{spec} \ a \) is contained in the set of roots. Let \( \psi \) be an irreducible (over \( \mathbf{K} \)) factor of the polynomial \( \varphi \). Because \( \varphi \) is minimal, the element \( \psi(a) \) is non-invertible. Consequently \( \psi(\psi(M)) = 0 \) for some \( M \) and therefore \( \psi \) is linear and its root belongs to \( \text{spec} \ a \). \( \square \)

In particular, if \( a \) is idempotent, i.e. \( a^2 = a \), and \( a \) is different from zero and the identity then \( \text{spec} \ a = \{0, 1\} \).

Corollary 4 (Spectral Mapping Theorem). For any polynomial \( \theta \) the identity \( \text{spec} \ \theta(a) = \theta(\text{spec} \ a) \) holds.

This \( \theta(a) \) is invertible if and only if \( \theta(\lambda) \neq 0 \) (\( \lambda \in \text{spec} \ a \)). Consequently we have the functional calculus \( \rho \mapsto \rho(a) \) for all rational functions \( \rho \) which are regular on \( \text{spec} \ a \) and \( \text{spec} \ \rho(a) = \rho(\text{spec} \ a) \).

For a maximal ideal \( M \) the natural homomorphism \( \chi_M: \mathfrak{A} \to \mathfrak{A}/M = \mathbf{K} \) is a character, i.e. a linear functional on \( \mathfrak{A} \) which is multiplicative in the sense that \( \chi_M(ab) = \chi_M(a)\chi_M(b) \) (\( \chi_M \neq 0 \) and in fact \( \chi_M(e) = 1 \)). The mapping \( M \mapsto \chi_M \) of the space of maximal ideals to the set of characters is bijective.

Let us consider some specific examples of spectral algebras.

Example 1. An algebra \( \mathfrak{A} \) over the field \( \mathbb{F}_2 = \{0, 1\} \) is said to be Boolean if all its elements are idempotent, i.e. the identity \( x^2 = x \) is satisfied in \( \mathfrak{A} \). Any Boolean algebra \( \mathfrak{A} \) is spectral since \( \mathfrak{A}/M = \mathbb{K} \) is a Boolean field for any maximal ideal \( M \) and then \( \mathfrak{A}/M = \mathbb{F}_2 \) since the identity element is the unique invertible idempotent.

We now observe that any Boolean algebra \( \mathfrak{A} \) is semisimple since, if \( a \in \text{Rad} \mathfrak{A} \), we have \( \text{spec} \ a = \{0\} \) so that \( a - e \) is invertible and then from the identity \( a(a - e) = 0 \) it follows that \( a = 0 \).

The image \( \mathfrak{B} \) of the algebra \( \mathfrak{A} \) under the canonical functional representation is isomorphic to \( \mathfrak{A} \). However the functions \( a \in \mathfrak{A} \) take only the values 0 and 1 and can therefore be identified with the corresponding subsets of the space \( \mathfrak{M} \) of maximal ideals. Moreover a sum transforms into a symmetric difference while a product becomes an intersection. Thus we obtain the result that any Boolean algebra is isomorphic to a certain algebra of sets (\textit{theorem of Stone}).

Example 2. Let \( \mathbf{K} = \mathbb{F}_q \) be any finite field (card \( \mathbf{K} = q = p^m \), where \( p \) is prime and \( m \geq 1 \)). The identity \( x^q = x \) holds in \( \mathbf{K} \). If it holds in a field \( E \supset \mathbf{K} \) then \( E = \mathbf{K} \) since the equation \( x^q = x \) cannot have more than \( q \) roots in \( E \). Therefore if this identity holds in an algebra \( \mathfrak{A} \) over \( \mathbf{K} \) then \( \mathfrak{A} \) is spectral. Also it is semisimple

\(^{19}\)These concepts are defined just as in the case of a linear operator. An algebraic operator on the space \( E \) is an algebraic element of the algebra \( L(E) \) and its spectrum in the previous sense if the minimal polynomial of the operator factorises into linear factors over the field \( \mathbf{K} \).
for the same reason as in the case \( q = 2 \). Therefore \( \mathfrak{U} \) is isomorphic to some subalgebra \( \Phi(\mathfrak{M}, F_q) \) (\( \mathfrak{M} \) is the space of maximal ideals).

**Example 3.** If the field \( K \) is algebraically closed, any algebra \( \mathfrak{U} \) over \( K \) which satisfies the condition

\[
\dim \mathfrak{U} < \text{card } K \tag{10}
\]
is spectral. In fact, let \( M \) be a maximal ideal, i.e. \( E = \mathfrak{U}/M \) is a field. If \( x \in E \setminus K \) then \( x - \lambda \) is invertible for all \( \lambda \in K \). Since the set of elements of the form \( (x - \lambda)^{-1} \) \( (\lambda \in K) \) has power \( \text{card } K \), it is linearly dependent. Hence it follows that \( x \) is algebraic over \( K \) and so \( x \in K \), which is a contradiction.

Condition (10) is satisfied if the algebra \( \mathfrak{U} \) has an infinite family \( \Gamma \) of generators such that \( \text{card } \Gamma < \text{card } K \); consequently such an \( \mathfrak{U} \) is spectral. This conclusion remains valid for finitely generated algebras but here a completely different argument is required; this uses the well-known *Nullstellensatz of Hilbert*. We can now generalise the last result in the following way.

Let us consider the algebra \( K[X] \) of polynomials in any collection \( X = \{x_i\}_{i \in I} \) of indeterminates over the field \( K \). It is clear that each polynomial \( f \) depends only on some finite collection of indeterminates. For any \( t \in K^I \) the value \( f(t) \in K \) is defined naturally and if \( f(t) = 0 \) then \( t \) is called a root of the polynomial \( f \). A common root of all polynomials belonging to a certain set is called a root of the set (without loss of generality we can take the set to be an ideal in \( K[X] \)).

**Theorem of Lang.** If the field \( K \) is algebraically closed and

\[
\text{card } X < \text{card } K \tag{11}
\]
then any ideal \( J \neq K[X] \) has a root.

**Proof.** If \( X \) is finite, condition (11) is automatically satisfied and we come back to Hilbert's *Nullstellensatz*. Suppose \( X \) is infinite.

We can assume that \( J = M \), a maximal ideal. The required root is \( t = (t_i)_{i \in I} \) where \( t_i \) is the image of the polynomial \( x_i \) in the field \( K[X]/M = K \).

**Corollary.** Suppose that the field \( K \) is algebraically closed and that (11) is satisfied. For consistency of a system of algebraic equations of the form

\[
f_v(t) = 0 \quad (f_v \in K[X], v \in N)
\]
in the unknown \( t = (t_i)_{i \in I} \) (\( N \) is some index set), it is necessary (trivial) and sufficient that each of its finite subsystems be consistent.

In fact, the ideal \( J \subset K[X] \) which is generated by the system \( (f_v)_{v \in N} \) is different from \( K[X] \). In the contrary case \( 1 \in J \), i.e. there exists a collection \( f_{v_1}, \ldots, f_{v_m} \) such that

\[
\sum_{j=1}^{m} f_{v_j} g_j = 1
\]
for some \( g_1, \ldots, g_m \) in \( K[X] \); but then the system of equations \( f_{v_j}(t) = 0 \) \( (1 \leq j \leq m) \) is not consistent.
Condition (11) is necessary as the following example shows:

\[ t_\omega (t_\omega - \alpha) = 1 \quad (\alpha \in K, \ X \supseteq \{x_\omega\}_{\omega \in X} \cup \{x_\omega\}) \]

where \( \omega \) is a fixed index which does not belong to \( K \).

Let us now consider the function algebra \( \Phi(S, K) \) for arbitrary \( S, K \). Associated with each point \( s \in S \) there is a character \( \varphi \mapsto \varphi(s) \) and correspondingly there is a maximal ideal \( M_s = \{ \varphi: \varphi(s) = 0 \} \). Clearly \( \Phi(S, K)/M_s = K \). The maximal ideals \( M_s \) are said to be standard. If \( S \) is finite all maximal ideals are standard and the algebra \( \Phi(S, K) \) is spectral.

Suppose that \( S \) is infinite. Let us consider the ideal \( \Phi_0(S, K) \) of functions with finite support. Every maximal ideal which contains it is non-standard and, conversely, any non-standard maximal ideal \( M \) contains \( \Phi_0(S, K) \) since \( M \not\supseteq M_s \) for any \( s \in S \), i.e. there exists \( \psi_s \in M \) such that \( \psi_s(s) = 1 \), and then \( \delta_s = \psi_s \delta_s \in M \).

**Theorem.** If \( S \) is infinite and \( \text{card } S \leq \text{card } K \) then for any non-standard maximal ideal \( M \subset \Phi(S, K) \) the field \( E = \Phi(S, K)/M \) is infinite-dimensional over \( K \).

**Proof.** We take an injective function \( \varepsilon \in \Phi(S, K) \) and show that the images of its powers in \( E \) are linearly independent. Suppose in contradiction to this that \( \pi(\varepsilon) \in M \) for some polynomial \( \pi \neq 0 \). The set of roots of the equation \( \pi(\varepsilon(s)) = 0 \) is finite and so its characteristic function \( \chi \) has finite support. Consequently \( \chi \in M \) and \( \pi(\varepsilon) + \chi \in M \). But \( \pi(\varepsilon(s)) + \chi(s) \neq 0 \) for all \( s \), i.e. \( \pi(\varepsilon) + \chi \) is invertible. \( \Box \)

**Corollary 1.** If \( S \) and \( K \) are infinite then \( \Phi(S, K) \) is not spectral.

A proof is required only in the case \( \text{card } S > \text{card } K \). But then there exists \( R \subset S \) such that \( \text{card } R = \text{card } K \) and, since \( R \) is infinite, the algebra \( \Phi(R, K) \) is not spectral. But \( \Phi(R, K) \approx \Phi(S, K)/J \), where \( J \) is the ideal of functions which vanish on \( R \). Moreover a factor algebra of any spectral algebra is also spectral. This follows from the general result: if \( h: \mathfrak{A} \rightarrow \mathfrak{B} \) is a surjective algebra homomorphism then the preimages of the maximal ideals of the algebra \( \mathfrak{B} \) are all maximal ideals of the algebra \( \mathfrak{A} \) which contain \( \text{Ker } h \); further \( \mathfrak{A}/\text{Ker } h \) is an isomorphic image of \( \mathfrak{B} \). We remark incidentally that if the field \( K \) is finite, \( K = \mathbb{F}_q \), then the algebra \( \Phi(S, K) \) is spectral since the identity \( \varphi^q = \varphi \) is satisfied in it.

**Corollary 2.** If \( S = \mathbb{N} = \{0, 1, 2, \ldots\} \) and \( K \) is any infinite field then for any non-standard maximal ideal \( M \subset \Phi(S, K) \) the field \( E = \Phi(S, K)/M \) is infinite-dimensional over \( K \).

The field \( E \) is said to be an ultrapower of the field \( K \). The ultrapowers of the field \( \mathbb{R} \) are in fact those of its extensions which are introduced in non-standard analysis by means of ultrafilters. This approach is equivalent to the previous one because of the 1-1 correspondence between the maximal ideals \( M \subset \Phi(S, K) \) and the ultrafilters \( U \) on \( S \): \( M(U) = \{ \varphi: \exists A \in U \text{ s.t. } \varphi|_A = 0 \} \).

\( ^5 \) In this context it is interesting to note that we can define the standard field \( \mathbb{R} \) as the factor algebra of the algebra of fundamental sequences over the field \( \mathbb{Q} \) of rational numbers by the maximal ideal of sequences tending to zero.
Among the function algebras on a topological space $S$ the algebra $C(S)$ of continuous functions (real or complex) first attracts attention. If $S$ is compact, all maximal ideals are standard (and consequently $C(S)$ is spectral). In fact, if the maximal ideal $M$ is such that for any point $s \in S$ there exists a function $\varphi_s \in M$ with $\varphi_s(s) \neq 0$, then $\varphi_s(t) \neq 0$ in some neighbourhood $V_s$ of the point $s$. By choosing a finite covering $V_{s_1}, \ldots, V_{s_n}$ we obtain the function

$$\varphi = \sum_{k=1}^{n} |\varphi_{s_k}|^2 = \sum_{k=1}^{n} \varphi_{s_k} \bar{\varphi}_{s_k} \in M$$

which has no zero, i.e. it is invertible in $C(S)$; this is a contradiction. Thus if $S$ is compact, the space of maximal ideals of the algebra $C(S)$ is naturally identified with $S$. In the non-compact situation we usually impose some condition at infinity on the functions; for example, we require boundedness, i.e. in place of the too extensive algebra $C(S)$ we consider the algebra $CB(S)$. If $S$ is completely regular (but in general not compact) the space of maximal ideals of the algebra $CB(S)$ is naturally identified with the Stone-Čech compactification $\beta(S)$ and $CB(S)$ is identified with $C(\beta(S))$ (I.M. Gelfand – A.N. Kolmogorov, 1939).

If $S$ is compact, then $C(S)$ is a Banach algebra under its usual norm. Algebras of continuous functions on compact spaces are important not only because they occur in analysis but also because of their special role in the general theory of commutative Banach algebras. We remark here that a commutative Banach algebra over the field of complex numbers is spectral (Gelfand – Mazur theorem).

In the Wiener algebra $W$ all maximal ideals are standard. Thus if a function $f \in W$ has no zero then it is invertible in $W$.

§ 2. Convex Analysis

From now on the ground field $K$ will always be $\mathbb{R}$ or $\mathbb{C}$ and where possible we will not distinguish these cases. As before the underlying space will be denoted by $E$.

2.1. Convex Sets. A set $X \subset E$ is said to be convex if for each pair of points $x_1, x_2 \in X$ the interval

$$[x_1, x_2] = \{x : x = \tau x_1 + (1 - \tau)x_2 \ (0 \leq \tau \leq 1)\}$$

is contained in $X$. Any subspace is convex.

Any point of the form

$$x = \sum_{k=1}^{n} \tau_k x_k \quad \left( \sum_{k=1}^{n} \tau_k = 1, \tau_k \geq 0 \right)$$

is said to be a convex combination of the points $x_1, \ldots, x_n$.

\[ \text{[The analogous assertion over } \mathbb{R} \text{ is false. For example, the Banach algebra } C \text{ over } \mathbb{R} \text{ is not spectral. See the proof of the Gelfand-Mazur theorem in Section 4.17.]} \]
Theorem of Carathéodory. If $E$ is finite-dimensional, $\dim E = r$, then any convex combination of the form (12) is a convex combination of some $r + 1$ points from the collection $(x_k)_{k=1}^n$.

In order that a set $X \subseteq E$ be convex it is necessary and sufficient that for each finite collection $(x_k)_{k=1}^n$ of points of $X$ all its convex combinations belong to $X$.

For any set $X \subseteq E$ the set of all convex combinations of all possible finite collections $(x_k)_{k=1}^n$ of points of $X$ is convex. This is the smallest convex set containing $X$. It is called the convex hull of the set $X$ and is denoted by $\text{Co } X$. If we note that the intersection of any family of convex sets is convex, we can define $\text{Co } X$ equivalently as the intersection of all convex sets containing $X$.

The convex hull of a finite collection $(x_k)_{k=1}^n$ is called a convex polytope. For $n = 1$ this is a point and for $n = 2$ it is an interval.

We note further that the sum of any family of convex sets is convex. For example, all affine manifolds $x + L$ ($x$ is a point and $L$ is a subspace) are convex.

A point $x$ is said to be an algebraic interior point of the convex set $X$ if $x \in X$ and for each $v \in E$ there exists $\varepsilon > 0$ such that $[x, x + \varepsilon v] \subseteq X$. The set of algebraic interior points of a set $X$ is called its algebraic interior and is denoted by $\text{ai } X$. It is easy to see that $\text{ai } X$ is convex. If $x \in \text{ai } X$ and $y \in X$ then the half-open interval $[x, y] = [x, y) \setminus \{y\}$ is contained in $\text{ai } X$. For fixed $x \in \text{ai } X$ the union of the half-open intervals $[x, y)$ ($y \in X$) coincides with $\text{ai } X$.

A convex set $X$ is said to be algebraically open if all its points are algebraic interior points, i.e. $X = \text{ai } X$. For any convex set its algebraic interior is algebraically open.

A convex set $X$ is said to be algebraically solid if its algebraic interior is non-empty.

A set $X \subseteq E$ is said to be circled if it is invariant with respect to the action $x \mapsto \zeta x$ of the group $|\zeta| = 1$. In the real case this means central symmetry, i.e. $x \in X \Rightarrow -x \in X$.

A set $X \subseteq E$ is said to be balanced if it is invariant with respect to the action $x \mapsto \rho x$ of the semigroup $|\rho| \leq 1$. For a convex set $X$ this property is equivalent to being circled.

A balanced convex set is said to be absolutely convex. Any subspace is absolutely convex.

Any point of the form

$$x = \sum_{k=1}^n \rho_k x_k \quad \left(\sum_{k=1}^n |\rho_k| \leq 1\right)$$

is called an absolutely convex combination of the points $x_1, x_2, \ldots, x_n$. In order that a set $X$ be absolutely convex it is necessary and sufficient that for each finite collection $(x_k)_{k=1}^n$ of points of $X$ all of its absolutely convex combinations belong to $X$.

The intersection of any family of absolutely convex sets is absolutely convex. If $X$ is absolutely convex then ai $X$ is also absolutely convex.

The absolutely convex hull of an arbitrary set $X \subseteq E$ is defined (in two equivalent ways) similarly to the convex hull.
A set $X \subset E$ is said to be conal if it is invariant with respect to the action $x \mapsto \lambda x$ of the group $\lambda > 0$. A conal convex set is called a wedge. Any subspace is a wedge. In order that a conal set be a wedge, it is necessary and sufficient that it also be an additive semigroup (i.e. invariant with respect to addition).

Any point of the form

$$x = \sum_{k=1}^{n} \lambda_k x_k \quad \left( \lambda_k \geq 0, \sum_{k=1}^{n} \lambda_k > 0 \right)$$

is called a non-negative combination of the points $x_1, \ldots, x_n$. In order that a set $X$ be a wedge it is necessary and sufficient that for each finite collection $(x_k)_{k}^n$ of points of $X$ all its non-negative combinations belong to $X$.

The intersection of any family of wedges is a wedge.

For any set $X \subset E$ the set of all non-negative combinations of all possible collections $(x_k)_{k}^n$ of points of $X$ is the smallest wedge which contains $X$, i.e. it coincides with the intersection of all wedges containing $X$.

If $K$ is a wedge, the intersection $\bigcap_{|k|=1} \zeta K$, if non-empty, is the largest subspace contained in $K$. It is called the edge of the wedge. A wedge is said to be a cone if its edge is zero or empty. The closed ray $\{y: y = \lambda x (\lambda \geq 0)\}$ and the open ray $\{y: y = \lambda x (\lambda > 0)\}$ defined by a non-zero vector $x$ are cones. In the first case the edge is zero, in the second it is empty.

For any wedge $K \neq E$ its algebraic interior $\text{ai} K$ is a cone (its edge is empty). A wedge $K$ is said to be generating if

$$\sum_{|k|=1} \zeta K = E.$$

Convex sets of one type or another arise naturally in connection with linear inequalities. Let $F = (f_i)_{i \in I}$ be some system of real linear functionals on $E$. Then the set of solutions of the system of linear inequalities

$$f_i(x) \geq \eta_i \quad (i \in I),$$

where the $\eta_i$ are arbitrary real numbers, is convex and the set of solutions of the corresponding homogeneous system

$$f_i(x) \geq 0 \quad (i \in I)$$

is a wedge. Its edge is the space of solutions of the homogeneous system of linear equations $f_i(x) = 0 (i \in I)$. The wedge of solutions of system (13) is therefore a cone if and only if the system $F$ is total. For the strict analogue of system (13), $f_i(x) > 0 (i \in I)$, the set of solutions is a cone\footnote{It is sufficient that just one of the inequalities be strict.} (it may be empty).

### 2.2. Convex Functionals

A real functional $q$ on $E$ is said to be subadditive if the triangle inequality

$$q(x + y) \leq q(x) + q(y) \quad (x, y \in E)$$

\footnote{Or on an additive subsemigroup $Q \subset E$.}
holds for it. This property implies the other inequalities:

\[ q(x - y) \geq q(x) - q(y); \quad 0 \leq q(0) \leq q(x) + q(-x). \]

It follows from the latter that \( \max(q(x), q(-x)) \geq 0 \).

The set of subadditive functionals contains all non-negative constants and all real additive functionals. A non-negative combination of subadditive functionals is subadditive, i.e. the subadditive functionals form a wedge. The least upper bound of any family of subadditive functionals which is pointwise bounded above is subadditive. A further useful construction is the composition \( \omega \circ q \), where \( \omega: \mathbb{R}_+ \to \mathbb{R}_+ \) is a non-decreasing subadditive function (for example, \( \omega(\xi) = \xi/(1 + \xi) \) \( \xi \geq 0 \)) and \( q \) is a non-negative subadditive functional; clearly under these conditions the functional \( \omega \circ q \) is subadditive.

A subadditive functional \( q \) on \( E \) is said to be sublinear if it is positively homogeneous, i.e. \( q(\lambda x) = \lambda q(x) \) for all \( \lambda > 0 \), \( x \in E \). In this case \( q(0) = 0 \), \( q(ax) \leq aq(x) \) for all real \( a \) and

\[
q\left(\sum_{k=1}^{n} \lambda_k x_k\right) \leq \sum_{k=1}^{n} \lambda_k q(x_k) \quad (\lambda_k \geq 0).
\]

The set of sublinear functionals is a wedge. The least upper bound of any family of sublinear functionals which is pointwise bounded above is sublinear.

**Example.** Let \( \{f_i\}_{i \in I} \) be a family of real linear functionals which is pointwise bounded above. Then the functional \( q(x) = \sup_i f_i(x) \) is sublinear.

A real functional \( g \) on \( E \) is said to be convex if

\[
g\left(\sum_{k=1}^{n} \tau_k x_k\right) \leq \sum_{k=1}^{n} \tau_k g(x_k) \quad \left(\tau_k \geq 0, \sum_{k=1}^{n} \tau_k = 1\right)
\]

(it is sufficient for this to be satisfied for \( n = 2 \)).

All sublinear functionals are convex but the class of convex functionals is considerably more extensive. For example, if \( \omega: \mathbb{R} \to \mathbb{R} \) is any non-decreasing convex function\(^5\), then for any convex functional \( q \) the composition \( \omega \circ q \) is convex. The set of convex functionals is a wedge. The least upper bound of any family of convex functionals which is pointwise bounded above is convex.

Each convex functional \( q \) generates two families of convex sets:

\[
\{x: q(x) \leq \alpha\}, \quad \{x: q(x) < \alpha\} \quad (\alpha \in \mathbb{R}).
\]

A set \( X \subset E \) is said to be absorbent if for each \( x \in E \) there exists \( \lambda > 0 \) such that \( \lambda^{-1} x \in X \). In other words \( E = \bigcup_{\lambda > 0} \lambda X \). Clearly any absorbent set contains zero. If \( X \) is convex a necessary and sufficient condition for absorbency is that \( 0 \in ai X \).

\(^5\) For this it is necessary and sufficient that \( \omega \) be the integral of a non-negative, non-decreasing function.
Let $X$ be a convex absorbent set. Since in this case $\lambda_1 X \subset \lambda_2 X (0 < \lambda_1 < \lambda_2)$, it is natural to introduce for $X$ its gauge (or Minkowski functional):

$$q_X(x) = \inf\{\lambda : \lambda > 0, x \in \lambda X\} \quad (x \in E).$$

This is a non-negative sublinear functional which 'almost' defines the set $X: q_X(x) < 1 \iff x \in \text{int} X$. Therefore, if $X$ is algebraically open, we have $X = \{x : q_X(x) < 1\}$. In general $q_X(x) \leq 1$ for $x \in X$. Conversely, for any non-negative sublinear functional $\varphi$, the set $D_\varphi = \{x : \varphi(x) < 1\}$ is convex, absorbent and algebraically open and its gauge coincides with $\varphi$. The set $D_\varphi$ is called the $\varphi$-unit ball.

We have a 1-1 correspondence $q \mapsto D_q$ between the non-negative sublinear functionals and the convex absorbent algebraically open sets.

If we remove the requirement of being algebraically open the collection of gauges is not increased since $q_X = q_{\text{int} X}$.

### 2.3. Seminorms and Norms.

A non-negative sublinear functional $p$ with the property that $p(\zeta x) = p(x)$ if $|\zeta| = 1$ is called a seminorm. In other words, the following are the defining properties of seminorms:

1) $p(x) \geq 0$;
2) $p(x + y) \leq p(x) + p(y)$;
3) $p(\alpha x) = |\alpha| p(x)$.

The last property is called absolute homogeneity.

**Example 1.** If $g$ is a sublinear functional then $p(x) = \sup_{|\zeta| = 1} g(\zeta x)$ is a seminorm.

**Example 2.** If $f$ is a linear functional then $|f|$ is a seminorm.

The gauge of any absolutely convex absorbent set is a seminorm. The mapping $p \mapsto D_p$ is a bijection of the set of seminorms onto the set of absolutely convex, absorbent, algebraically open sets. This property allows us to introduce seminorms in a purely geometric way.

Associated with each seminorm $p$ is the subspace $\text{Ker} p = \{x : p(x) = 0\}$, its kernel. Seminorms with zero kernel are norms. In other words, norms are those seminorms whose unit spheres contain no rays.

It follows from the inequality

$$|p(x) - p(y)| \leq p(x - y)$$

that if $x - y \in \text{Ker} p$ then $p(x) = p(y)$. Thus in a natural way $p$ generates a seminorm on the factor space $E/\text{Ker} p$. This factor seminorm is a norm.

The set of all seminorms (norms) is a cone. The least upper bound of any family of seminorms which is pointwise bounded above is a seminorm.

**Example.** Let $F = \{f_i\}_{i \in I}$ be a pointwise bounded family of linear functionals. Then the functional

$$p(x) = \sup_i |f_i(x)|$$

is a seminorm.

---

$^6$ $q_X(x) = 0 (x \neq 0)$ if and only if the ray defined by the vector $x$ is contained in the set $X$. 
is a seminorm. We will see in what follows that this is a universal method for constructing seminorms. In order that the seminorm \((14)\) be a norm, it is necessary and sufficient that the family \(F\) be total. In the general case

\[
\text{Ker } p = \bigcap_i \text{Ker } f_i.
\]

2.4. The Hahn-Banach Theorem. Let \(q\) be a sublinear functional on \(E\) and \(f\) a real linear functional on a subspace \(L \subseteq E\) which satisfies the condition

\[ f(x) \leq q(x) \quad (x \in L). \]

Then \(f\) can be extended to the whole space \(E\) in such a way that the majorant \(q\) is preserved, i.e. there exists a real linear functional \(\tilde{f}\) on \(E\) such that \(\tilde{f}|_L = f\) and \(\tilde{f}(x) \leq q(x)\) \((x \in E)\).

The proof is carried out with the help of Zorn’s lemma. The basic step consists of extending by one dimension, i.e. to a subspace \(\tilde{L} = L + \text{Lin}_R \{e\}\) where \(e \notin L\). It is sufficient to choose \(\tilde{f}(e)\) so that \(f(x) + \alpha \tilde{f}(e) \leq q(x + \alpha e)\) for all \(\alpha \in \mathbb{R}, x \in L\).

But it is easy to see that

\[
\sup_{x \in L, \alpha > 0} \frac{f(x) - q(x - \alpha e)}{\alpha} \leq \inf_{x \in L, \beta > 0} \frac{q(y + \beta e) - f(y)}{\beta}.
\]

Any value of \(\tilde{f}(e)\) lying between these bounds satisfies the required condition.

This theorem is one of the basic principles of linear functional analysis.

**Corollary 1.** Let \(p\) be a seminorm on \(E\) and \(f\) any linear functional on a subspace \(L \subseteq E\) which satisfies the condition \(|f(x)| \leq p(x)\) \((x \in L)\). Then there exists a linear functional \(\tilde{f}\) on \(E\) such that \(\tilde{f}|_L = f\) and \(|\tilde{f}(x)| \leq p(x)\) \((x \in E)\).

**Proof.** If \(E\) is a real space we can extend \(f\) in such a way that \(\tilde{f}(x) \leq p(x)\) \((x \in E)\). Substituting \(-x\) for \(x\) here, we obtain the required result. In the complex case it is sufficient to extend the real linear functional \(g = \text{re } f\) so that \(\tilde{g}(x) \leq p(x)\) \((x \in E)\) and then to put \(\tilde{f}(x) = \tilde{g}(x) - i\tilde{g}(ix)\).

A linear functional \(f\) on \(E\) is said to be **subordinate** to a seminorm \(p\) if there exists \(c \geq 0\) such that \(|f(x)| \leq cp(x)\) \((x \in E)\). The greatest lower bound of such \(c\) is called the **\(p\)-norm** of the functional \(f\) and is denoted by \(\|f\|_p\). Thus

\[
|f(x)| \leq \|f\|_p p(x) \quad (x \in E)
\]

and it is impossible to substitute a smaller constant in place of the \(p\)-norm. Clearly if \(p \neq 0\)

\[
\|f\|_p = \sup_{p(x) \neq 0} \frac{|f(x)|}{p(x)} = \sup_{p(x) = 1} |f(x)| = \sup_{p(x) \leq 1} |f(x)| = \sup_{p(x) < 1} |f(x)|.
\]

The set of linear functionals subordinate to a fixed seminorm \(p\) is a linear space and moreover \(\|\cdot\|_p\) is a norm on this space.

**Corollary 2.** Let \(p\) be a seminorm on \(E\) and let \(f\) be any linear functional on a subspace \(L \subseteq E\) which is subordinate to \(p\). Then \(f\) can be extended to the whole of \(E\) without change of \(p\)-norm, i.e. there exists a linear functional \(\tilde{f}\) on \(E\) which is subordinate to \(p\) and such that \(\tilde{f}|_L = f\) and \(\|\tilde{f}\|_p = \|f\|_p\).
This follows from Corollary 1 on replacing $p$ by $Cp$, where $C = \|f\|_p$.

**Corollary 3.** Let $p$ be a seminorm on $E$, let $x_0 \in E$ and suppose that $p(x_0) \neq 0$. Then there exists a linear functional $f$ which is subordinate to $p$ and such that $f(x_0) = p(x_0)$ and $\|f\|_p = 1$.

Any such functional is said to be $p$-supporting for the vector $x_0$. It can be constructed by defining the linear functional $f_0(ax_0) = ap(x_0)$ on the one-dimensional subspace $\text{Lin}\{x_0\}$ and then extending $f_0$ to $E$ without change of $p$-norm.

Let us now denote by $S_p^p$ the set of linear functionals which are subordinate to $p$ and such that $\|f\|_p = 1$. Then, as a result of Corollary 3,

$$p(x) = \max_{f \in S_p^p} |f(x)| = \max_{f \in S_p^p} \Re f(x).$$

Thus any seminorm is the maximum of some family of linear functionals (necessarily subordinate to it).

The following more general assertion is established similarly to Corollary 3: if $q$ is a non-negative sublinear functional on $E$, then for each vector $x_0$ there exists a $q$-supporting real linear functional $f$:

$$f(x_0) = q(x_0), \quad f(x) \leq q(x) \quad (x \in E).$$

### 2.5. Separating Hyperplanes.

From here up to the end of Section 2 the basic space $E$ is assumed to be real.

Let us consider an arbitrary hyperplane $H = \{x : f(x) = \alpha\}$ ($f$ is a non-zero linear functional, $\alpha \in \mathbb{R}$). The complement $E \setminus H$ is divided into two algebraically open half-spaces $H_+ = \{x : f(x) > \alpha\}, H_- = \{x : f(x) < \alpha\}$.

We also put $\overline{H}_+ = H_+ \cup H = E \setminus H_-$ and $\overline{H}_- = H_- \cup H = E \setminus H_+$. All these sets are convex.

For any two non-intersecting sets $X, Y \subset E$ we say that a hyperplane $H$ (or a linear functional $f$ which defines it) separates $X$ and $Y$ if one of these sets lies in $H_-$ and the other in $\overline{H}_+$ or one lies in $H_+$ and the other in $\overline{H}_-$. Clearly this is equivalent to the assertion that $H$ separates the convex hulls $\text{Co} \ X$ and $\text{Co} \ Y$; it is therefore enough to study the separation of convex sets. The fundamental results in this direction are due to Ascoli (1932), Mazur (1933) and Eidelheit (1936).

**Theorem.** For any non-empty convex algebraically open set $X$ and any non-empty convex set $Y$ disjoint from $X$ there exists a separating hyperplane.

**Proof.** Let $x_0 \in X, y_0 \in Y$ and put $z_0 = y_0 - x_0$. Let us consider the convex algebraically open set $D = X - Y + z_0$. Since $0 \in aiD$ it follows that $D$ is absorbent. Since $X \cap Y = \emptyset$ we have that $z_0 \notin D$ and so $q_D(z_0) \geq 1$. We construct a linear functional $f$ which is $q_D$-supporting at the point $z_0$. Then if $x \in X$,
$y \in Y$ we have $f(x - y + z_0) < 1$ and hence $f(x) < f(y)$ and $\alpha \equiv \sup_{x \in X} f(x) \leq \inf_{y \in Y} f(y)$. Since $X$ is algebraically open, the upper bound $\alpha$ is not attained. If we now take the hyperplane $H = \{v: f(v) = \alpha\}$ then $X \subset H_-$ and $Y \subset H_+$. □

We shall say that a hyperplane $H$ (or a linear functional $f$ which defines it) almost separates the sets $X, Y \subset E$ if one of them lies in $H_-$ and other in $H_+$ (moreover $X$ and $Y$ can even intersect). Here we can also restrict attention to convex sets. From the theorem on the separating hyperplane follows

**Corollary 1.** Let $X$ and $Y$ be convex sets with $X$ algebraically solid and $Y$ non-empty and disjoint from $\text{ai} \ X$. Then there exists a hyperplane which almost separates $X$ and $Y$.

In fact a hyperplane which separates $\text{ai} \ X$ and $Y$ almost separates $X$ and $Y$.

**Remark.** For any two non-intersecting non-empty convex sets in a finite-dimensional space there exists a hyperplane which almost separates them.

The fundamental result, which is generally known as the geometric form of the Hahn-Banach theorem, follows in turn from Corollary 1.

**Corollary 2.** Let $X$ be a convex algebraically solid set and $A$ an affine manifold such that $A \cap \text{ai} \ X = \emptyset$. Then there exists a hyperplane $H \supset A$ such that $H \cap \text{ai} \ X = \emptyset$ (i.e. $\text{ai} \ X \subset H_-$ or $\text{ai} \ X \subset H_+$).

**Proof.** Let $G = \{v: f(v) = \alpha\}$ be a hyperplane which almost separates $X$ and $A$: $f(x) \leq \alpha \ (x \in X), \ f(y) \geq \alpha \ (y \in A)$. If $A = x_0 + L$, where $L$ is a subspace, we have $f(z) \geq \alpha - f(x_0) \ (z \in L)$ and so $f(z) = 0 \ (z \in L)$. Consequently, the hyperplane $H = \{y: f(y) = f(x_0)\}$ contains the manifold $A$. Moreover $f(x) \leq f(x_0) \ (x \in X)$ and hence $f(x) < f(x_0) \ (x \in \text{ai} \ X)$, i.e. $\text{ai} \ X \subset H_-$. □

**Corollary 3.** Any convex algebraically open set $X \neq E$ is the intersection of some family of algebraically open half-spaces.

In fact, for each point $v \notin X$ there exists a hyperplane $H(v)$ such that $v \in H(v)$ and $X \subset H_-(v)$. Since moreover $v \notin H_-(v)$ we have

$$X = \bigcap_{v \notin X} H_-(v).$$

In other words, $X$ coincides with the set of solutions of a system of linear inequalities of the form $f_v(x) < \alpha_v \ (v \notin X)$.

Now let $X$ be a convex algebraically solid set, $X \neq E$. Then from the above

$$X \subset \bigcap_{v \notin \text{ai} X} H_-(v),$$

i.e. $X$ is contained in the set of solutions of a system of linear inequalities of the form $f_v(x) \leq \alpha_v \ (v \notin \text{ai} \ X)$. 

A hyperplane $H = \{u: f(u) = \alpha\}$ is said to support a convex set $X$ if $\sup_{x \in X} f(x) = \alpha$.

**Corollary 4.** If $X$ is a convex algebraically solid set different from $E$, then

$$X \subset \bigcap_{H \in S} \overline{H},$$

where $S$ is the family of support hyperplanes of $X$.

### 2.6. Non-negative Linear Functionals.

Associated with each wedge $K$ is the wedge $^9 K^*$ of linear functionals which are $K$-non-negative in the sense that $f(x) \geq 0$ ($x \in K$) (notation: $f \geq 0$). All functionals $f \geq 0$ ($f \neq 0$) are positive on the algebraic interior of the wedge $K$. In the case where $K$ is a cone, a functional $f$ is said to be $K$-positive ($f > 0$) if $f(x) > 0$ for all $x \in K$, $x \neq 0$.

**Theorem of M.G. Krejn.** Suppose that the subspace $L \subset E$ contains an algebraic interior point $\hat{x}$ of the wedge $^1 E$. Then any $(K \cap L)$-non-negative linear functional $f$ on $L$ can be extended to a $K$-non-negative linear functional on $E$.

**Proof.** If $f \neq 0$ then Ker $f \cap ai K = \emptyset$. Therefore Ker $f$ is contained in some hyperplane $H = \{x: g(x) = 0\}$ such that $H \cap ai X = \emptyset$. The functional $g$ provides the required extension. □

**Corollary.** For any algebraically solid wedge $K \neq E$ there exists a $K$-non-negative linear functional $g \neq 0$.

**Proof.** If $\hat{x} \in ai K$ then $(-\hat{x}) \notin K$ since $K \neq E$. On the line $L = Lin\{\hat{x}\}$ there is the $(K \cap L)$-non-negative linear functional $f \neq 0$: $f(\alpha \hat{x}) = \alpha$. It remains to extend it to a $K$-non-negative functional on $E$. □

**Example 1.** Let $S$ be a space with a finite measure $\mu$, let $E = B(S, R)$, the space of bounded functions $S \rightarrow R$ and let $K = B_{+}(S, R)$, the cone of non-negative functions. Its algebraic interior $ai$ $K$ consists of those functions with positive lower bound. Let us consider the subspace $L = L^1(S, \mu)$ of summable functions and the $(K \cap L)$-non-negative linear functional

$$J(\varphi) = \int_S \varphi \, d\mu$$
on it. Since $1 \in L \cap ai K$, the functional $J$ can be extended to a $K$-non-negative linear functional $\tilde{J}$ on $B(S, R)$. Now for any set $M \subset S$ we can put $\mu(M) = J(\chi_M)$, where $\chi_M$ is its characteristic function. With this we have constructed a finitely additive extension of the measure $\mu$ to the algebra of all subsets of the set $S$. Correspondingly,

---

8 As above, this term carries over to functionals. We note that if $q$ is a non-negative sublinear functional then a functional which is $q$-supporting at the point $x_0$ ($q(x_0) \neq 0$) is a support functional for the set $\{x: q(x) \leq q(x_0)\}$.

9 A cone if the wedge $K$ is generating.

10 This wedge must therefore be algebraically solid.
where the integral can be interpreted as before, i.e. as the least upper bound of Lebesgue integral sums. Naturally, this integral is only finitely additive.

From what has been said it is also clear that the integral with respect to a finite, finitely additive measure gives the general form of a non-negative linear functional on $B(S, \mathbb{R})$.

For the space $C(S)$ of continuous functions on a compact space $S$ we have the theorem of F. Riesz on the general form of a non-negative linear functional:

$$f(\varphi) = \int_X \varphi \, d\mu,$$

where $\mu$ is a regular Borel measure\(^\text{11}\) which is uniquely determined by the functional $f$. The first step of the proof is the extension of $f$ to $B(S)$.

**Example 2.** Many variants of the classical moment problem are to be found in the following abstract scheme. In the space $E$ there are given a wedge $K$ ($0 \in K$) and some system $U = \{u_i\}_{i \in I}$ of vectors such that $\text{Lin } U \cap ai K \neq \emptyset$. It is required to find a $K$-non-negative linear functional $f$ which satisfies the system of equations

$$f(u_i) = c_i \quad (i \in I),$$

where the $c_i$ are given scalars (the moments of the required functional $f$ with respect to the system $U$). It turns out that for solvability of this abstract moment problem it is necessary and sufficient that the implication

$$\sum_{i \in I} \alpha_i u_i \in K \Rightarrow \sum_{i \in I} \alpha_i c_i \geq 0$$

should hold for systems of scalars $(\alpha_i)_{i \in I}$ with finite support. We only need to establish the sufficiency. Without loss of generality we can assume that the system $U$ is linearly independent. Then $f$ is uniquely defined on $L = \text{Lin } U$ by the given equations and from the condition we see that it is $(K \cap L)$-non-negative; finally we can extend it to a $K$-non-negative linear functional on the whole of $E$.

\(^{11}\) A Borel measure is said to be regular if its value on each Borel set $M$ coincides with the supremum of its values on the compact sets $Q \subseteq M$ and with the infimum of its values on the open sets $G \supset M$. In what follows, all measures on locally compact topological spaces are assumed to be regular Borel measures (or this property is to be proved).

The problem of constructing a linear functional $f$ which is subordinate to a given seminorm $p$ and satisfies a system of equations $f(u_i) = c_i$ ($i \in I$) can be solved similarly with the help of the Hahn-Banach theorem. A criterion for solvability is that the inequality

$$\left| \sum_{i \in I} \alpha_i c_i \right| \leq C p \left( \sum_{i \in I} \alpha_i u_i \right) \quad (C \text{ is a positive constant})$$

should hold for all $(\alpha_i)_{i \in I}$ of finite support. Moreover the solution will satisfy the condition $\|f\|_p \leq C$.\ 
In particular let us consider from this point of view the trigonometric moment problem
\[ \int_{-\pi}^{\pi} e^{inx} d\sigma(x) = c_n \quad (n = 0, \pm 1, \pm 2, \ldots). \]

Here we are looking for a non-negative linear functional
\[ f(\varphi) = \int_{-\pi}^{\pi} \varphi(x) d\sigma(x) \]
on \( C(-\pi, \pi) \) which satisfies the equations \( f(e_n) = c_n \) \( (e_n(x) = e^{inx}, n = 0, \pm 1, \pm 2, \ldots) \). A criterion for solvability is that, for each non-negative trigonometric polynomial \( T(x) = \sum_n \alpha_n e^{inx} \), the combination \( \sum_n \alpha_n c_n \) must be non-negative. But, by the well-known theorem of Fejér and F. Riesz, the condition \( T(x) \geq 0 \) is equivalent to \( T(x) = |V(x)|^2 \) where \( V \) is a trigonometric polynomial.

The Riesz-Herglotz criterion (positivity of the matrix \( (c_{n-m})_n,m=0 \)) follows easily from what has been said.

2.7. Ordered Linear Spaces. This class of spaces was introduced and first studied by L.V. Kantorovich (1935). A linear space \( E \) is said to be ordered if there is given on it a partial ordering which is compatible with its linear structure:

1) \( x < y \Rightarrow x + z < y + z \); 2) \( x < y \Rightarrow ax < ay \) \( (a > 0) \).

It follows from property 1) that \( (x < y) \land (z < w) \Rightarrow x + z < y + w \) and moreover \( x < y \Rightarrow (-y) \leq (-x) \), so that, by (2), \( x < y \Rightarrow ay \leq ax \) \( (a < 0) \).

The standard method of ordering a linear space is to choose some cone \( K \) containing zero and then to define
\[ x \leq y \Leftrightarrow y - x \in K; \]
in particular \( y \geq 0 \Leftrightarrow y \in K \), so that \( x \leq y \Leftrightarrow y - x \geq 0 \). Conversely, if the space \( E \) is already ordered, the set \( K = \{ y : y \geq 0 \} \) is a cone containing zero and \( x \leq y \Leftrightarrow y - x \in K \).

The elements of the cone \( K \) are said to be non-negative in accordance with their characterisation by the inequality \( y \geq 0 \).

Example. The linear space \( \Phi(S, \mathbb{R}) \) of all real functions on any set \( S \) is ordered pointwise: \( \varphi \leq \psi \Leftrightarrow \varphi(s) \leq \psi(s) \) \( (s \in S) \). This ordering is defined by the cone of non-negative functions.

Any subspace \( L \) of an ordered linear space \( E \) is automatically ordered, the corresponding cone being \( K \cap L \) where \( K \) is the cone in \( E \).

We now describe the standard method of ordering the space of homomorphisms \( \text{Hom}(E, E_1) \), where \( E_1 \) is also an ordered linear space and \( Q \subseteq E_1 \) is the corresponding cone. We assume that the cone \( K \) in \( E \) is generating. Then the homomorphisms \( h : E \rightarrow E_1 \), which map \( K \) into \( Q \) form a cone in \( \text{Hom}(E, E_1) \) and the corresponding order is defined by this. In particular, \( h \geq 0 \) means that \( x \geq 0 \Rightarrow hx \geq 0 \) and, because of the linearity of the mapping \( h \), this is equivalent to its monotonicity: \( x \leq y \Rightarrow hx \leq hy \).
Taking $E_1 = \mathbb{R}(Q = \mathbb{R}_+)$ we obtain the natural ordering on the space of linear functionals on $E$. It is defined by the cone $K^*$.

We note further here that the definitions of subadditivity, sublinearity and convexity used in Section 2.2 for real functionals carry over word for word to mappings of any linear space into any ordered linear space.

Of special interest among ordered linear spaces are lattices, in which (by definition) for each pair of vectors $x$, $y$ there exists a supremum $\text{sup}(x, y)$, their least upper bound$^{13}$:

$$\text{sup}(x, y) = \min\{z: z \geq x, z \geq y\}.$$  

We note immediately that$^{14}$

$$\text{sup}(x + z, y + z) = \text{sup}(x, y) + z, \quad \text{sup}(\lambda x, \lambda y) = \lambda \text{sup}(x, y) \quad (\lambda > 0).$$

**Example.** The space $\Phi(S, \mathbb{R})$ is a lattice:

$$\text{(sup}(\varphi, \psi))(s) = \max(\varphi(s), \psi(s)).$$

If an ordered linear space is a lattice it is called a vector (or linear$^{15}$) lattice or a Riesz space.

A subspace $L$ of a vector lattice is called a sublattice if $\text{sup}(x, y) \in L$ for all $x, y \in L$. For example, $\Phi_0(S, \mathbb{R})$ is a sublattice of $\Phi(S, \mathbb{R})$.

A cone $K$ which orders a space $E$ is said to be minihedral if $\text{sup}(x, y)$ exists for all $x, y \in K$. Clearly the cone of a vector lattice is minihedral. Moreover it is generating since if $x_+ = \text{sup}(x, 0)$ and $x_- = \text{sup}(-x, 0)$ then $x_+ \geq 0, x_- \geq 0$ and $x = x_+ - x_-.$

**Remark 1.** If $x = u - v$ where $u \geq 0$ and $v \geq 0$, then $u = x_+ + z$ and $v = x_- + z$ where $z \geq 0$. In this sense, among the stated representations of a vector $x$ as the difference of two non-negative vectors, there is a minimal one: $x = x_+ - x_-.$

**Remark 2.** Any minihedral generating cone $K \subset E$ turns $E$ into a lattice. For the proof we put $x - y = u - v (u, v \in K)$ and then $\text{sup}(x, y) = \text{sup}(u, v) + (y - v)$.

Now we introduce the modulus of an element $x$ of a vector lattice $E$ by putting $|x| = x_+ + x_- = \text{sup}(x, -x)$. The mapping $x \mapsto |x|$ is sublinear$^{16}$ and moreover $|x| > 0 (x \neq 0), |0| = 0$, i.e. the usual properties of modulus are preserved under this generalisation.

Any element $r \geq 0$ of a vector lattice $E$ with cone $K$ defines a cube $C_r = \{x: |x| \leq r\} = \{x: -r \leq x \leq r\}$. A linear functional $f$ on $E$ is said to be locally

$^{13}$The existence of an infimum $\text{inf}(x, y)$, the greatest lower bound (also part of the definition of a lattice), is automatically guaranteed here: $\text{inf}(x, y) = -\text{sup}(-x, -y)$.

$^{14}$In particular the relation $\text{sup}(x, y) + \text{inf}(x, y) = x + y$ follows from this.

$^{15}$In fact these words are often omitted.

$^{16}$This follows from the subadditivity of the mappings $x \mapsto x_+, x \mapsto x_-$. We note also that $|\alpha x| = |\alpha||x|$ for all $\alpha \in \mathbb{R}, x \in E.$
§3. Linear Topology

**bounded** if it is bounded on each cube. These functionals form a linear space \( E' \) which is ordered by means of the cone of non-negative functionals (we note that if \( f \geq 0 \) then \( f \in E' \)).

**Theorem 1.** In order that \( f \) be locally bounded it is necessary and sufficient that it be the difference of two non-negative functionals.

The sufficiency is trivial. For the proof of the necessity we put \( g(x) = \sup_{0 \leq y \leq x} f(y) \) \( (x \geq 0) \). This functional is additive and positive-homogeneous on the cone \( K \). It is uniquely extended by linearity to \( K - K \) and then to the whole of \( E \). By construction, \( g(x) \geq 0 \) and \( g(x) \geq f(x) \) \( (x \geq 0) \). It remains to put \( h(x) = g(x) - f(x) \). □

The constructed functional \( g \) is \( \sup(f, 0) \). At the same time we have also proved

**Theorem 2.** The space \( E' \) of locally bounded linear functionals on a vector lattice \( E \) is a lattice.

A vector lattice \( E \) is said to be **Dedekind complete** if any non-empty subset \( X \subset E \) which is bounded above has a supremum \( \sup X \). If this is only required for countable sets, the lattice will be called **Dedekind \( \sigma \)-complete**. The space \( \Phi(S, \mathbb{R}) \) is a Dedekind complete lattice but \( \Phi_0(S, \mathbb{R}) \) is not even Dedekind \( \sigma \)-complete.

Making Theorem 2 more precise, we can also assert that \( E' \) is a Dedekind complete lattice.

§3. Linear Topology

**3.1. Linear Topological Spaces.** If on a linear space \( E \) there is given a Hausdorff topology with respect to which the operations of addition and multiplication by a scalar are continuous\(^1\), then \( E \) is called a **linear topological space** (LTS). In particular, under addition an LTS is an abelian topological group. Therefore the system of open (closed) sets in an LTS is invariant with respect to translations \( x \mapsto x + y \) and **homotheties** \( x \mapsto ax \) \( (a \neq 0) \) (i.e. these are homeomorphisms). We note also that affine properties such as convexity, being circled or balanced and absolute convexity are invariant with respect to closure.

A linear topology on \( E \) is completely determined by a base of neighbourhoods of zero (which clearly has to be assigned so that not only general topological requirements but also linear topological ones are satisfied). We note that in any LTS there exists a base of balanced neighbourhoods of zero. In fact if \( U \) is some neighbourhood of zero, then, by the continuity of multiplication by a scalar, there exist \( \varepsilon > 0 \) and a neighbourhood \( V \) of zero such that \( \alpha V \subset U \) if \( |\alpha| < \varepsilon \). Then

\(^1\) i.e. the mappings \((x, y) \mapsto x + y\) and \((x, \alpha) \mapsto ax\) of \( E \times E \) and \( K \times E \) \( (K = \mathbb{R} \text{ or } C) \) under their product topologies into \( E \) are continuous.
\[ W = \bigcup_{|x| < \varepsilon} \alpha V \]
is a balanced neighbourhood of zero contained in \( U \).

In any LTS all open sets are algebraically open.

We note further that any neighbourhood of zero in an LTS is an absorbent set.

We give several simple example of LTSs.

**Example 1.** The usual topology on the ground field \( K \) is linear.

**Example 2.** The function space \( \Phi(S, K) \) is an LTS under the topology of pointwise convergence (i.e. the product topology since \( \Phi(S, K) = K^S \)).

Since any subspace of an LTS is an LTS under the induced topology, we have in particular that \( \Phi_0(S, K) \) is an LTS. This provides a method of converting an arbitrary linear space \( E \) into an LTS since \( E = \Phi_0(B, K) \) where \( B \) is a basis.

Generally speaking the topology obtained depends on the choice of basis. However, the following result holds.

**Theorem.** On any finite-dimensional space (and only on such a space) there is a unique linear topology.

We note that a topological criterion for finite-dimensionality of an LTS is local compactness.

A metric is one of the basic methods of defining a topology. If on a linear space \( E \) there is given a metric \( d \) which is invariant with respect to translations\(^2\) and satisfies the conditions

\[
\text{d}(\alpha x, 0) \leq d(x, 0) \quad (|\alpha| \leq 1), \quad \lim_{\varepsilon \to 0} d(\varepsilon x, 0) = 0 \quad (15)
\]

then \( E \), equipped with this metric, is an LTS. The function \( v(x) = d(x, 0) \) is a pseudonorm, i.e. 1) \( v(x) > 0 \) \( (x \neq 0) \), \( v(0) = 0 \), 2) \( v(\alpha x) \leq v(x) \) \( (|\alpha| \leq 1) \), 3) \( v(x + y) \leq v(x) + v(y) \), and it is affinely continuous at zero, i.e. continuous in each direction\(^3\): \( \lim_{\varepsilon \to 0} v(\varepsilon x) = v(0) \) \( (=0) \). Conversely, any pseudonorm which is affinely continuous at zero determines an invariant metric \( d(x, y) = v(x - y) \) which satisfies the conditions \( (15) \), i.e. finally, it generates a linear topology. In this construction are included not only normed spaces but also certain others of definite interest.

**Example 1.** In the space \( (S, \mu) \) of measurable functions on a set \( S \) with finite measure \( \mu \) the pseudonorm

\[
v(\varphi) = \int_S \frac{|\varphi(s)|}{1 + |\varphi(s)|} \, d\mu
\]
is affinely continuous at zero. The metric produced by it defines the linear topology of convergence in measure.

\(^2\)E. such that \( d(x + z, y + z) = d(x, y) \) (in what follows it is simply called *invariant*).

\(^3\)Continuity in each direction at a given point is clearly a concept which does not require a topology on \( E \).
Example 2. The space $L^p(S, \mu) (0 < p < 1)$ is defined by the condition

$$v(\varphi) = \int_S |\varphi|^p \, d\mu < \infty$$

and it is topologised by this pseudonorm.

Remark. Sometimes a linear topology is introduced by means of a quasinorm, i.e. a functional $N$ which possesses all the properties of a norm except possibly the triangle inequality, which is replaced by $N(x + y) \leq C[N(x) + N(y)]$ where $C$ is a positive constant. A base of neighbourhoods of zero is given by inequalities of the type $N(x) < \varepsilon$ ($\varepsilon > 0$).

If the linear topology on $E$ is given explicitly by means of a metric then $E$ is called a linear metric space; but if it is only a question of the existence of such a metric on an LTS, then it is said to be metrizable. Clearly, satisfying the first axiom of countability, i.e. the existence of a countable base of neighbourhoods of zero, is a necessary condition for this.

Birkhoff-Kakutani Theorem. If an LTS satisfies the first axiom of countability then it is metrizable and moreover the metric which defines its topology can be chosen to be invariant and to satisfy conditions (15).

Remark. The authors mentioned in fact resolved the problem of metrizability for topological groups. It was shown additionally that on any (Hausdorff) topological group the bounded uniformly continuous functions separate the points. Hence it follows that the underlying space of a topological group is completely regular. In particular any LTS is completely regular. The weaker property of regularity is obtained from simple considerations: if $G$ is a topological group, $X \subset G$ is a closed set, $x \notin X$ and $U$ is a symmetric neighbourhood of zero such that $U + U$ is contained in $(G \setminus X) - x$, then the neighbourhood $x + U$ of the point $x$ does not intersect the neighbourhood $X + U$ of the set $X$.

From the Birkhoff-Kakutani theorem follows

Corollary. In any linear metric space there exists a topologically equivalent invariant metric satisfying condition (15).

LTSs $E_1$, $E_2$ are said to be topologically isomorphic if there exists a linear homeomorphic mapping (topological isomorphism) of $E_1 \to E_2$. Isometric isomorphism of linear normed spaces is a special case of this (i.e. the existence of a linear isometry from $E_1$ onto $E_2$).

If $E_1$, $E_2$ are finite-dimensional then topological isomorphism is equivalent to algebraic isomorphism, i.e. to the equality $\dim E_1 = \dim E_2$. The problem of classifying infinite-dimensional LTSs up to topological isomorphism is extremely complicated. Certain advances in this direction are connected with the gener-

---

4As before, functions which are equal almost everywhere are identified.
5The space $\mathcal{C}(S, \mathbb{K})$ with uncountable $S$ is not metrizable.
6This word may be omitted in a context where it is naturally implied.
alisation of the concept of dimension (Banach-Mazur, 1933; A.N. Kolmogorov, 1958). In his time Banach introduced the conjecture that all infinite-dimensional separable Banach spaces are homeomorphic (not linearly, in general). M.I. Kadets proved the validity of this conjecture in 1966. However, for example, the spaces \( L^p(0, 1) \) \((p > 1)\) are mutually non-isomorphic (Banach-Mazur, 1933). The spaces of continuous functions on compact metric spaces with cardinality of the continuum are isomorphic to each other (A.A. Milyutin, 1966).

A metrizable LTS \( E \) is said to be complete, or to be a Fréchet space, if it is complete (in the usual sense) with respect to some (and consequently any) invariant metric which generates the topology on \( E \).

**Remark 1.** Completeness may be lost if the invariance of the metric is removed. For example, the usual topology on \( \mathbb{R} \) is generated by any metric of the type \( d(x, y) = |\theta(x) - \theta(y)| \), where \( \theta: \mathbb{R} \to \mathbb{R} \) is a local homeomorphism, although \( \mathbb{R} \) is not complete with respect to this metric if the function \( \theta \) is bounded.

**Remark 2.** In the class of all LTSs we can define the property of completeness by requiring the convergence of every fundamental net \( (x_i) \) (i.e. a net \( (x_i) \) such that \( \lim_{i,k}(x_i - x_k) = 0 \)). For metrizable LTSs this definition reduces to the previous one.

**Remark 3.** We can complete any LTS in the standard way. Moreover the completion of a metrizable LTS will be a Fréchet space, the completion of a normed space will be a Banach space and the completion of a pre-Hilbert space will be a Hilbert space.

Fréchet spaces have the important property of barrelledness. A **barrelled space** is an LTS in which each barrel contains a neighbourhood of zero, where **barrel** means an absorbent absolutely convex closed set. The barrelledness of a Fréchet space follows from Baire's theorem that a complete metric space is of the second category, i.e. it cannot be the countable union of nowhere dense sets. In fact if an LTS \( E \) is of second category then it is barrelled. For the proof we consider an arbitrary barrel \( T \subseteq E \). Since

\[
E = \bigcup_{n=1}^{\infty} nT
\]

and all the \( nT \) are closed, for some \( m \) the set \( mT \) is not nowhere dense, i.e. it has an interior point; then \( T \) also has this property. But if \( x \in \text{int } T \), then \( -x \in \text{int } T \) and consequently \( 0 \in \text{int } T \); i.e. a neighbourhood of zero is contained in \( T \).

We can characterise barrelledness in terms of seminorms. In any LTS continuity of a seminorm \( p \) is equivalent to its continuity at zero and for this in turn it is necessary and sufficient that the \( p \)-unit ball \( D_p \) should contain a neighbourhood of zero. For an LTS to be barrelled it is necessary and sufficient that each lower

\footnote{This also applies to any subadditive functional \( q \) with \( q(0) = 0 \).}

\footnote{The last criterion is correct not only for seminorms but also for any non-negative sublinear functionals. This can also be expressed in the form of a boundedness requirement on some neighbourhood of zero.}
semicontinuous seminorm on it be continuous. In fact, let \( E \) be barrelled and let \( p \) be a lower semicontinuous seminorm. Then the set \( \{ x : p(x) \leq 1 \} \) is closed and consequently it is a barrel. If \( U \) is a neighbourhood of zero which is contained in it, then \( U \subset D_p \) since all points of \( U \) are algebraic interior points. Conversely, if \( T \) is a barrel then its gauge \( q_T \) is a lower semicontinuous seminorm and then, by the condition, it is continuous and consequently the \( q_T \)-unit ball is a neighbourhood of zero.

**Banach-Steinhaus Theorem.** If a family of continuous seminorms \( \{ p_i \}_{i \in I} \) in a barrelled space \( E \) is pointwise bounded, then \( p(x) = \sup_i p_i(x) \) is a continuous seminorm.

In fact the seminorm \( p \) is lower semicontinuous.

Under the conditions of the Banach-Steinhaus theorem the family of seminorms \( \{ p_i \}_{i \in I} \) turns out to be bounded on some neighbourhood of zero (in the case of a Banach space the family is bounded on each ball), and so the theorem is often called the principle of uniform boundedness (in the case of a Banach space we obtain \( p_i(x) \leq C \| x \| \) \((C = \text{constant}, i \in I)\) or, what is equivalent, boundedness of norms for the family \( \{ p_i \}_{i \in I} \). Uniform boundedness of a family of seminorms in any LTS is equivalent to its equicontinuity.

From the point of view of classical analysis the Banach-Steinhaus theorem is the quintessence of the method of condensation of singularities. We give some examples of its application.

**Example 1.** We show that there exists a continuous function \( f(t) \) with period \( 2\pi \), for which the partial sums \( S_n(t; f) \) of its Fourier series are not uniformly bounded. Assuming the contrary, we would have in the Banach space of continuous functions on the circle a sequence of seminorms \( p_n(f) = \max_{|t| \leq \pi} |S_n(t; f)| \) \((n = 0, 1, 2, \ldots)\) which is pointwise bounded. By the Banach-Steinhaus theorem, \( p_n(f) \leq C \| f \| \) \((C = \text{constant})\). But then the Lebesgue constants \( L_n \) turn out to be bounded which, as is well-known, is false.

**Example 2.** Many methods for the summation of numerical series are covered by the following scheme proposed by Toeplitz (1911). Let a triangular numerical matrix \( A = (\lambda_{ik})_{0 \leq i \leq k} \) be given. We associate with any series having partial sums \( s_k \) \((k = 0, 1, 2, \ldots)\) the sequence

\[
\sigma_i = \sum_{k=0}^{i} \lambda_{ik} s_k \quad (i = 0, 1, 2, \ldots),
\]

and if the latter converges to \( \sigma \), then we shall say that the initial series is summable to \( \sigma \) by means of the matrix \( A \). The question arises of the regularity of such methods of summation. It turns out that for regularity it is necessary and sufficient that the following conditions be satisfied:

1) \( \lim_{i \to \infty} \lambda_{ik} = 0 \); 2) \( \lim_{i \to \infty} \sum_{k=0}^{i} \lambda_{ik} = 1 \); 3) \( \sup_{i} \sum_{k=0}^{i} |\lambda_{ik}| < \infty \).

We shall prove only the necessity of 3). Let us consider the Banach space \( c \) of
convergent sequences \( s = (s_n)_{n \geq 0} \) with norm \( \|s\| = \sup_n |s_n| \). Under the condition, the sequence of continuous seminorms \( |\sigma_i(s)| (i = 0, 1, 2, \ldots) \) converges pointwise and consequently it is pointwise bounded in \( c \). By the Banach-Steinhaus theorem

\[
\left| \sum_{k=0}^i \lambda_{ik} s_k \right| \leq C \|s\| \quad (C = \text{constant}; i = 0, 1, 2, \ldots).
\]

It just remains for us to put \( s_k = \text{sgn} \lambda_{ik} (0 \leq k \leq i) \) for arbitrary \( i \).

**Example 3.** Let us consider on a compact space \( X \) with finite measure \( \mu \) the quadrature process

\[
J_n(f) = \sum_{j=0}^n \rho_{nj} f(x_{nj})
\]

with nodes \( \{x_{nj}\}_{0 \leq j \leq n} \) and coefficients \( \{\rho_{nj}\}_{0 \leq j \leq n} \) which satisfy the condition

\[
J_n(u_k) = \int_X u_k \, d\mu \quad (0 \leq k \leq n),
\]

where \( \{u_k\}_{n \geq 0} \) is a complete Markov system. If the process converges for all \( f \in C(X) \) then, by the Banach-Steinhaus theorem,

\[
\left| \sum_{j=0}^n \rho_{nj} f(x_{nj}) \right| \leq C \|f\| \quad (C = \text{constant}; n = 0, 1, 2, \ldots).
\]

Since for each \( n \) there exists in \( C(X) \) a function \( f \) satisfying the conditions \( f(x_{nj}) = \text{sgn} \rho_{nj} (0 \leq j \leq n) \), \( \|f\| \leq 1 \), it follows that

\[
\sum_{j=0}^n |\rho_{nj}| \leq C.
\]

It is easy to show that this necessary condition for convergence is also sufficient.

Returning to general problems, let us assume that a total family \( \{p_i\}_{i \in I} \) of seminorms is given on the linear space \( E \): thus for any \( x \neq 0 \) there exists \( i \) such that \( p_i(x) \neq 0 \). Then we can introduce on \( E \) a linear topology with base of neighbourhoods of zero

\[
U_{i_1, i_2, \ldots, i_m; \epsilon} = \left\{ x : \max_{1 \leq k \leq m} p_{i_k}(x) < \epsilon \right\} \quad (\epsilon > 0)^9.
\]

All these neighbourhoods are convex (and even absolutely convex).

A linear topology which has a base of convex neighbourhoods of zero is said to be locally convex and an LTS with such a topology is called a locally convex space (LCS). The credit for separating out this class of LTSs, which has become firmly established in contemporary mathematics, belongs to von Neumann (1935).

Specialising the previous example let us consider a total family \( \{f_i\}_{i \in I} \) of linear functionals on a linear space \( E \). The base of neighbourhoods of zero

\[
9^9 This is the weakest topology under which all the seminorms \( p_i \) are continuous.
defines a locally convex topology on $E$.

**Example 1.** The topology of pointwise convergence in the space $\Phi(S, K)$ is defined by the family of linear functionals $f_\ell(\phi) = \phi(s)(s \in S)$ and consequently it is locally convex. In this topology $\Phi(S, K)$ is complete.

**Example 2.** Let $S$ be any Hausdorff topological space and let $\phi \in C(S)$. For each compact set $Q \subset S$ we put $p_Q(\phi) = \max_Q |\phi|$. The topology of uniform convergence on compact sets is defined by means of this family of seminorms on $C(S)$. In this topology $C(S)$ is complete whenever $S$ is locally compact.

We can define similarly on $C(S)$ a topology of uniform convergence on any family of compact sets $\{Q_i\}_{i \in I}$ which covers all of $S$. The topology of pointwise convergence is included in this scheme.

The topology of uniform convergence on any set $S$ can be defined in the space $B(S)$ by means of the norm $\|\phi\| = \sup_S |\phi|$, with respect to which $B(S)$ is a Banach space.

**Theorem.** In any LCS the topology is defined by a certain total family of seminorms.

We can form this family from the gauges of the absolutely convex neighbourhoods of zero (such neighbourhoods form a base at zero).

If the topology on $E$ is defined by a countable family of seminorms $\{p_n\}_0^\infty$ then $E$ is metrizable. A suitable invariant metric in this case has the form:

$$d(x, y) = \sum_{n=0}^{\infty} \frac{p_n(x - y)}{2^n \left(1 + p_n(x - y)\right)}.$$

Thus, for example, if $S$ is $\sigma$-compact (i.e. it is the union of countably many compact sets) then $C(S)$ is a locally convex Fréchet space (but not a Banach space if $S$ is not compact).

Let us pass to the consideration of subspaces in an LTS $E$. They are all LTSs under the induced topology (LCSs if $E$ is of this type). Particularly important are closed subspaces. For example, in the case of Fréchet (or Banach) spaces only closed subspaces are Fréchet (or Banach) spaces, since subspaces which are not closed are incomplete. Any closed subspace of a complete LTS is complete.

If $L$ is a closed subspace in any LTS $E$ then the factor space $E/L$ with its standard topology is an LTS. If in addition $E$ is an LCS then $E/L$ is also an LCS and if $E$ is barrelled then $E/L$ is also barrelled. If $E$ is metrizable and $d$ is an

---

10 This is the weakest amongst those for which all functionals $f_i$ are continuous. Any linear functional on $E$ which is continuous under this topology is a linear combination of the functionals $f_i (i \in I)$.

11 This is the strongest topology under which the natural homomorphism $j: E \to E/L$ is continuous, i.e. the open subsets of $E/L$ are precisely those sets whose preimages are open in $E$. We note that in addition the homomorphism $j$ is open. If the closedness of $L$ is dropped then $E/L$ fails to be Hausdorff.
invariant metric on $E$ then $E/L$ is metrizable by means of the invariant metric

$$d(X, Y) = \inf_{x, y} d(x, y),$$

(16)

where $x, y \in E$ run through the corresponding classes $X, Y \in E/L$. Further if $E$ is complete then $E/L$ is also complete in the metrizable case. Thus for Fréchet spaces all factor spaces by a closed subspace are Fréchet spaces. The situation is analogous for Banach spaces. In this case definition (16) reduces to the form

$$\|X\| = \inf_{x \in X} \|x\|.$$ 

A factor space of a Hilbert space by a closed subspace is a Hilbert space.

In any LTS $E$ the closure of any subspace is a closed subspace. It follows from this that any hyperplane in an LTS is either closed or dense.

We note further that finite-dimensional subspaces in any LTS are closed.

In any LTS a subset $X$ is said to be bounded if it is absorbed by each neighbourhood $U$ of zero ($U$ can be assumed to be balanced) i.e. $X \subset \lambda U$ for some $\lambda = \lambda(X, U) > 0$. For example, any finite subset and in general any compact subset is bounded.

The class of bounded sets is invariant with respect to finite unions, linear combinations and closure. For LCSs we can add to this list the formation of absolutely convex hulls. Each subset of a bounded set in an LTS is clearly bounded. In any LTS boundedness of a set $X$ is equivalent to the condition that if $\{x_n\}_{n=1}^\infty \subset X, \{a_n\}_{n=1}^\infty \subset \mathbb{R}$ and $\lim_{n \to \infty} a_n = 0$ then $\lim_{n \to \infty} a_n x_n = 0$. The boundedness of a set is therefore equivalent to the boundedness of all of its countable subsets.

It is clear that if the topology is weakened the collection of bounded sets is not diminished.

In order that a set in an LTS be bounded, it is necessary, and in the case of an LCS also sufficient, that all continuous seminorms (or even just the basic seminorms) be bounded on it. For example, in $\Phi(S, \mathbb{K})$ boundedness is equivalent to pointwise boundedness.

For boundedness of a set in a normed space it is necessary and sufficient that it be contained in some ball $\|x\| < \rho$. The converse of this obvious remark leads to a criterion for normability of an LTS, i.e. for the existence of a norm which generates the given topology.

**Theorem of Kolmogorov.** In order that an LTS be normable it is necessary and sufficient that there exist a bounded convex neighbourhood $U$ of zero.

In fact $U$ must then contain an absolutely convex neighbourhood $V$. The gauge of the neighbourhood $V$ is the desired norm.

An LTS is said to be locally bounded if it has a bounded neighbourhood $W$ of zero. In this case for any sequence of numbers $\alpha_n > 0$ which converges to zero the sets $\alpha_n W$ form a base of neighbourhoods of zero. Consequently any locally bounded LTS is metrizable and local boundedness of an LCS is equivalent to
normability. Therefore, if a metrizable LCS is not normable then the balls in it are not bounded. Thus the concept of boundedness in the linear topological sense does not coincide with the usual metric boundedness in the class of metrizable LTSs (even LCSs). In a normed space these concepts are equivalent and in a linear metric space with invariant metric topological boundedness implies metric boundedness.

3.2. Continuous Linear Functionals. Continuity of a linear functional \( f \neq 0 \) is equivalent to the closedness of the hyperplane \( H_\alpha = \{ x : f(x) = \alpha \} \) for any \( \alpha \). In fact closedness of \( H_\alpha \) for any \( \alpha \) is equivalent to closedness of \( H_0 \) and for closed \( H_0 \) the factor space \( E/H_0 \) is an LTS and \( f \) induces a bijection \( \tilde{f} : E/H_0 \to K \). Transferring the linear topology from \( E/H_0 \) to \( K \) with the aid of \( \tilde{f} \), we obtain from the commutative diagram

\[
\begin{array}{ccc}
E & \xrightarrow{f} & K \\
\downarrow j & & \downarrow j \\
E/H_0 & \xrightarrow{} & \\
\end{array}
\]

in which \( j \) is the natural homomorphism, that \( f \) is continuous. It just remains to take account of the uniqueness of the linear topology on \( K \).

Continuity of a linear functional on an LTS is equivalent to its continuity at zero and this in turn is equivalent to the boundedness of the functional on some neighbourhood of zero. In a normed space we can choose the open unit ball \( D = \{ x : \| x \| < 1 \} \) as such a neighbourhood of zero. We also consider the closed unit ball \( \bar{D} = \{ x : \| x \| \leq 1 \} \) and the unit sphere \( S = \bar{D} \setminus D = \{ x : \| x \| = 1 \} \). Boundedness of a linear functional \( f \) on any one of these three sets \( D, \bar{D}, S \) implies its boundedness on the other two and thereby is equivalent to its continuity. But if \( f \) is bounded on \( S \) (and only in this case) then it is subordinate to the basic norm \( \| \cdot \| \), and moreover its \( \| \cdot \| \)-norm (subsequently called simply the norm of \( f \)) is equal to

\[
\| f \| = \sup_{\| x \| = 1} |f(x)|.
\]

This least upper bound is not altered on replacing the unit sphere by either the open or the closed unit ball.

In view of what has been stated, linear functionals on a normed space, which are subordinate to the basic norm, are said to be bounded but in fact this property is equivalent to continuity. As a consequence of a corollary of the Hahn-Banach theorem the basic norm can itself be represented in the form

\[
\| x \| = \max_{\| f \| = 1} |f(x)|.
\]

The essential difference of this formula from the previous one, its 'dual', is that here the least upper bound is attained (at a functional which is supporting at \( x \)).
**Example.** In the Banach space $l^1$ of numerical sequences $x = (\xi_n)_{n=0}^\infty$ with norm

$$\|x\| = \sum_{n=0}^\infty |\xi_n|$$

any bounded sequence $f = (\alpha_n)_{n=0}^\infty$ defines a continuous linear functional:

$$f(x) = \sum_{n=0}^\infty \alpha_n \xi_n; \quad \|f\| = \sup_n |\alpha_n|$$

If the sequence $(|\alpha_n|)_{n=0}^\infty$ does not attain its least upper bound then $|f(x)|$ also does not attain its least upper bound on the unit sphere $\|x\| = 1$.

In any LCS the continuity of a linear functional is equivalent to its subordination to some continuous seminorm and is also equivalent to its subordination to some finite collection of basic seminorms.

For LCSs the Hahn-Banach theorem guarantees that it is possible to extend any continuous linear functional defined on a subspace to a continuous linear functional on the whole space. In the case of a normed space there is an extension which preserves the norm.

The geometric form of the Hahn-Banach theorem in any real LTS guarantees that if an affine manifold $A$ does not intersect the interior of a convex set $X$ with non-empty interior then it is contained in a closed hyperplane which does not intersect int $X$. In fact int $X$ is a non-empty convex, open, and consequently algebraically open, set. We take a hyperplane $H \ni A$ which does not intersect int $X$. It is closed since it is not everywhere dense.

The topological variant of the theorem on the separating hyperplane is established similarly: for any non-empty open convex set $X$ and any non-empty convex set $Y$ which does not intersect it there is a closed separating hyperplane.

A set $X$ in an LTS is called a convex body if it is convex, closed and has non-empty interior. At each boundary point of a convex body $X \neq E$ there exists a support hyperplane; all such hyperplanes are closed and $X$ is the intersection of all the closed half-spaces of the form $f(x) \leq \alpha$ which contain it, where the pairs $f, \alpha$ correspond to all possible support hyperplanes.

In an LCS there is a further variant of the theorem on the separating hyperplane: for any non-empty closed convex set $X$ and any non-empty compact convex set $Q$ which is disjoint from $X$ there exists a closed separating hyperplane. This situation is reduced to the previous one by constructing disjoint convex neighbourhoods for $X$ and $Q$. The hyperplane which we want turns out to be even strictly separating:

$$\sup_{x \in X} f(x) < \inf_{y \in Q} f(y).$$

In particular, if $L$ is a closed subspace in an LCS $E$ and $x_0 \notin L$, there exists a continuous linear functional $f$ on $E$ such that $f|_L = 0$, $f(x_0) \neq 0$ (we can take $f(x_0) = 1$). This conclusion also remains valid in the complex case. Thus if $L \neq E$ there exists a continuous linear functional $f \neq 0$ such that $f|_L = 0$. From this follows the very useful.
Lemma of F. Riesz. If \( L \) is a closed subspace of a normed linear space \( E \) and \( L \neq E \) then for any \( \delta \) (\( 0 < \delta < 1 \)) there exists in \( E \) a vector \( x \) (\( \| x \| = 1 \)) whose distance from \( L \) satisfies the inequality:

\[
d(x, L) = \inf_{y \in L} \| x - y \| \geq \delta.
\]

Proof. Let \( f \) be any linear functional such that \( \| f \| = 1, f|_L = 0 \). We choose \( x \) to satisfy the conditions \( \| x \| = 1 \), \( | f(x) | \geq \delta \). Then for all \( y \in L \) we have

\[
\| x - y \| \geq | f(x) - f(y) | = | f(x) | \geq \delta.
\]

In any LCS (real or complex) each closed subspace is the intersection of the closed hyperplanes which contain it, i.e. it is defined by a system of linear equations of the form \( f_i(x) = 0 \), where the \( f_i \) are continuous linear functionals. If \( E \) is real then any closed convex set in \( E \) is the intersection of the closed half-spaces which contain it, i.e. it is defined by a system of linear inequalities of the form \( f_i(x) \leq \alpha_i \), where the \( f_i \) are continuous linear functionals and the \( \alpha_i \) are real numbers. For compact convex sets we need consider only half-spaces corresponding to hyperplanes of support.

If \( K \) is a wedge with non-empty interior in a real LTS \( E \) then all \( K \)-non-negative linear functionals are positive on the interior of \( K \) and are consequently continuous. Thus in the topological variant of the theorem on the extension of non-negative functionals the resulting functionals are automatically continuous. For any wedge \( K \subseteq E \) the set of non-negative continuous linear functionals is denoted by \( K^* \). If \( K \) is a closed cone in a Banach space \( E \) then for each vector \( w \in K \) there exists a functional \( f \in K^* \) such that \( f(w) > 0 \) (theorem of M.G. Krejn).

For any LTS \( E \) (real or complex) the continuous linear functionals form a linear space \( E^* \), called the conjugate\(^{12} \) of \( E \). The case \( E^* = 0 \) is possible\(^{13} \), although if \( E \) is an LCS then \( E^* \) is total. We can use the latter property to introduce a locally convex topology on \( E \) defined by means of the set \( E^* \). Since the new topology is weaker than the initial topology (as a result of the continuity of all \( f \in E^* \)), it is called the weak topology (or w-topology) while the original topology is called the strong topology. We note that such a weakening of the topology does not change the set of continuous linear functionals, i.e. \( E^* \) remains as before since, if a linear functional \( f \) is weakly continuous, there exists \( \delta > 0 \) such that from a certain system of inequalities of the form \( | f_i(x) | < \delta \) (\( f_i \in E^*, 1 \leq i \leq n \)) follows the inequality \( | f(x) | < 1 \). But then it follows from the equations \( f_i(x) = 0 \) (\( 1 \leq i \leq n \)) that \( f(x) = 0 \). Thus \( f \in \text{Lin}\{f_1, \ldots, f_n\} \subseteq E^* \). It is also clear from what has been asserted that in general if the topology on \( E \) is given by a total space \( F \) of linear functionals then \( E^* = F \).

The question of topologies on \( E^* \) leads to a very general consideration which we will leave aside for the time being. However if \( E \) is a normed space then \( E^* \)

---

\(^{12}\)More precisely, its topological conjugate. Clearly \( E^* \subseteq E^* \) but in general \( E^* \neq E^* \); for example \( \Phi(S, K)^* = \Phi_0(S, K) \). If \( E \) is finite-dimensional then \( E^* = E^* \).

\(^{13}\)As an example we can consider \( E = L^p(0, 1) \) for \( 0 < p < 1 \). This is a Fréchet space in which there are no open convex subsets except \( E, \emptyset \).
is also normed (as was mentioned above) and even a Banach space (as is easily verified). We can now consider the second conjugate \( E^{**} \) and the canonical homomorphism \( v: E \to E^{**} \) defined by the formula

\[
(vx)(f) = f(x) \quad (x \in E, f \in E^*).
\]

Since \( |(vx)(f)| \leq \|f\| \|x\| \) and equality is attained at a support functional \( f \) for \( x \), then \( \|vx\| = \|x\| \), i.e. \( v \) is an isometry. Thus we can always consider a normed space \( E \) as a subspace of the Banach space \( E^{**} \) and identify the completion of the space \( E \) with its closure \( \overline{E} \subset E^{**} \) \( (E \) is a Banach space if and only if it is closed in \( E^{**} \).)

A Banach space \( E \) is said to be reflexive\(^{14} \) if \( E = E^{**} \) (i.e. if the canonical isometry \( v \) is surjective). In contrast to the algebraic situation there exist infinite-dimensional reflexive Banach spaces (for example, all Hilbert spaces are of this type (see below)). If \( E \) is finite-dimensional then it is reflexive since in this case all linear functionals are continuous.

Reflexivity of the space \( E^* \) is equivalent to the reflexivity of the space \( E \).

James (1964) established the following remarkable criterion: for the reflexivity of a Banach space \( E \) it is necessary and sufficient that each continuous linear functional on \( E \) should attain its norm on the unit sphere \(^{15} \). It is appropriate to note here that for any Banach space the set of continuous linear functionals which attain their norms is dense in \( E^* \) (Bishop-Phelps, 1961).

James (1951) constructed an example of a Banach space \( E \) for which \( \dim(E^{**}/E) = 1 \), and moreover \( E^{**} \) is isometric to \( E \) (but of course not canonically).

A Banach space is said to be uniformly convex if whenever \( 0 < \varepsilon < 2 \) there exists \( \delta > 0 \) such that from the conditions \( \|x\| = \|y\| = 1, \|x - y\| > \varepsilon \) there follows \( \|(x + y)/2\| < 1 - \delta \). For example, the spaces \( L^p(S, \mu) \) for \( 1 < p < \infty \) are of this type (Clarkson, 1936). Any Hilbert space is uniformly convex\(^{16} \). All uniformly convex spaces are reflexive (D.P. Milman, 1938). This classical result follows easily from James’ criterion.

A Banach space is said to be uniformly smooth if for any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that from the conditions \( \|x\| = 1, \|y\| < \delta \) there follows \( \|x + y\| + \|x - y\| < 2 + \varepsilon \|y\| \). Uniform smoothness (convexity) of the space \( E \) is equivalent to uniform convexity (smoothness) of the space \( E^* \) (V.L. Shmulian, 1940). Thus all uniformly smooth Banach spaces are reflexive.

All spaces \( L^p(S, \mu) \) \( (1 < p < \infty) \) are uniformly smooth. Any Hilbert space is uniformly smooth.

The unit sphere of a Banach space is said to be smooth at the point \( x \) if there is a unique support functional \( f_x \) at this point. A Banach space is said to be smooth

\(^{14} \) More precisely, topologically reflexive.

\(^{15} \) The necessity is obvious: for any \( f \in E^* \) there exists a support functional \( \zeta \in E^{**} \) \( (\zeta(f) = \|f\|, \|\zeta\| = 1) \) and, if \( E^{**} = E \), then \( \zeta = x \in E \), \( \zeta(f) = f(x), \|\zeta\| = \|x\| \).

\(^{16} \) As a consequence of the parallelogram identity: \( \|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + \|y\|^2 \). Incidentally, this identity for a Banach space implies that it is a Hilbert space (Jordan-von Neumann, 1935).
if its unit sphere is smooth at each point. In this case there is defined the support mapping \( x \mapsto f_x \) from the unit sphere of the space \( E \) into the unit sphere of the space \( E^* \).

Smoothness of the unit sphere at a point \( x \) is equivalent to differentiability of the norm at this point in the sense of Gâteaux, i.e. the existence of the limit

\[
\langle x, y \rangle = \lim_{\tau \to 0} \frac{\|x + \tau y\| - \|x\|}{\tau}
\]

for all \( y \). The space \( E \) is uniformly smooth if and only if this limit exists uniformly for all \( x, y (\|x\| = 1, \|y\| = 1) \).

In any separable Banach space the set of smooth points is dense in the unit sphere (Mazur, 1933).

A Banach space is said to be strictly convex if there are no line segments on its unit sphere or equivalently if the condition \( \|x + y\| = \|x\| + \|y\| \) (\( x, y \neq 0 \)) implies \( y = ax \) \((a > 0)\). Clearly this property is weaker than uniform convexity. Strict convexity of a reflexive space is equivalent to smoothness of the conjugate space. A more general result holds: if \( E^* \) is smooth or strictly convex then \( E \) is correspondingly strictly convex or smooth. The converse assertion is false (Day 1955; Klee, 1959).

Returning to the situation of an arbitrary LCS \( E \), we note that each closed convex set \( X \subset E \) is weakly closed (and consequently for any convex \( X \subset E \) its weak and strong closures coincide). In fact, if \( x_0 \notin X \), there exists a functional \( f \in E^* \) such that \( \sup_{x \in X, \tau \to x_0} \text{Re} \ f(x) < \text{Re} \ f(x_0) \), and then \( E \setminus X \) contains the weak neighbourhood \( \{y : \text{Re} \ f(y - x_0) < \varepsilon\} \) for sufficiently small \( \varepsilon > 0 \).

We remark further that in an LCS the weak topology coincides with the strong topology on compact sets since the weak topology is still Hausdorff.

In specific function spaces we can usually obtain an explicit description (the general form) of all continuous linear functionals. In this series of results, many of which were established long ago by F. Riesz, the first place is taken by his theorem on the general form of a continuous linear functional on the space \( C(S) \) of continuous (real or complex) functions on a compact space \( S \):

\[
f(\varphi) = \int_S \varphi \, d\omega,
\]

where \( \omega \) is a uniquely defined real or complex measure. Its proof makes use of the extension of a functional to \( B(S) \) by means of the Hahn-Banach theorem, which allows the possibility of applying the functional to characteristic functions of sets.
If we define the norm (total variation) of the measure $\omega$ by putting

$$\|\omega\| = \int_S |d\omega|,$$

then it can be proved that $\|f\| = \|\omega\|$. The space $M(S)$ of measures on $S$ (with the properties described above) turns out to be a Banach space which is isometrically isomorphic to the space $C(S)^*$. In what follows these two spaces are identified.

If $S$ is a locally compact space then continuous linear functionals on $C(S)$ have a form similar to the previous one but with measures of compact support (a measure is said to have compact support if it is equal to zero on sets which are disjoint from some compact set).

In the space $L^p(S, \mu)$ ($1 \leq p < \infty$) continuous linear functionals are identified with the elements of the space $L^q(S, \mu)$, where $q$ is related to $p$ by Young's relation $p^{-1} + q^{-1} = 1$:

$$f(\phi) = \int_S \phi \psi \, d\mu \quad (\phi \in L^p, \psi \in L^q).$$

In addition $\|f\| = \|\psi\|_{L^q}$ which allows us to consider $(L^p)^* = L^q$. However we must add that if $p = 1$ then $q = \infty$ and $L^{\infty}(S, \mu)$ is the space of measurable functions which are bounded except for values taken on a set of measure zero. As usual, two functions which coincide almost everywhere are identified. The norm on $L^{\infty}(S, \mu)$ is defined as

$$\text{ess.sup}|\psi| = \inf_{\mu(X) = 0} \sup_{x \in S \setminus X} |\psi(x)|$$

and under it $L^{\infty}(S, \mu)$ becomes a Banach space which is conjugate to $L^1(S, \mu)$.

From what has been said it is immediately clear that the spaces $L^p$ are reflexive for $1 < p < \infty$. The spaces $B(S), C(S), M(S), L^1(S, \mu), L^{\infty}(S, \mu)$ are not reflexive (except for finite $S$).

The case $p = 2$ ($q = 2$) is distinguished by the fact that the conjugate space can be identified with the initial space. This property is possessed by all Hilbert spaces (it is a stronger requirement than reflexivity).

**Theorem of F. Riesz.** If $E$ is a Hilbert space then the general form of a continuous linear functional is given by the formula

$$f(x) = (x, y) \quad (x \in E, y \in E);$$

further $\|f\| = \|y\|$.

**Remark.** For a Hilbert space $E$ the isometry of $E^*$ onto $E$ defined by the theorem of F. Riesz is linear in the real case but *semilinear* in the complex case (i.e. additive and $af \mapsto \bar{a}y(f)$).

3.3. Complete Systems and Topological Bases. We have already discussed on several occasions complete systems of functions in certain function spaces. In exactly the same way a system of vectors $V = (v_i)_{i \in I}$ in an LTS $E$ is said to be
complete if its linear hull is dense, i.e. its closed linear hull coincides with $E$. For example any Hamel basis is such a system but a complete system can be considerably thinner.

**Example.** The countable system of functions $t^n (0 \leq t \leq 1; n = 0, 1, 2, \ldots)$ is complete\(^{19}\) in $C[0, 1]$, whereas every Hamel basis has the power of the continuum.

If there exists in an LTS $E$ a countable complete system of vectors then $E$ is separable.

If $E$ is an LCS a necessary and sufficient condition for a system $V \subset E$ to be complete is that any continuous linear functional $f$ which annihilates $V$ be identically zero. Therefore in proofs that specific systems are complete, theorems on the general form of continuous linear functionals are usually applied.

**Example 1.** Let us consider in $L^2(-\pi, \pi)$ the system of exponentials $(e^{i\lambda_nt})_{n \in \mathbb{Z}}$ with arbitrary (in general complex) exponents $\lambda_n$. The continuous linear functionals which annihilate the given system are identified with those functions $\varphi \in L^2(-\pi, \pi)$ such that

$$\int_{-\pi}^{\pi} \varphi(t) e^{i\lambda_n t} dt = 0 \quad (n \in \mathbb{Z}).$$

Putting

$$\tilde{\varphi}(\lambda) = \int_{-\pi}^{\pi} \varphi(t) e^{i\lambda t} dt,$$

we arrive at the entire functions of class $B^2_{\pi}$ with roots at the points $\lambda_n$. Consequently, in order that the given system of exponentials be complete, it is necessary and sufficient that $\{\lambda_n\}$ be a uniqueness set for the interpolation problem $\psi(\lambda_n) = 0 \ (n \in \mathbb{Z})$ in the class $B^2_{\pi}$. The trivial situation of this type is where the set $\{\lambda_n\}$ has a limit point. Therefore we may consider that $\lambda_n \to \infty$. In this direction we obtain, for example, the following result (A.I. Khejlits, 1969): let $\lambda_n = n + \theta(n)$, where $\theta(n) = O(1)$; then the system $(e^{i\lambda_n t})$ is complete in $L^2(-\pi, \pi)$ if and only if

$$\sum_{m=-\infty}^{\infty} \exp(2 \mathrm{Re} \Delta_m(\theta)) = \infty,$$

where

$$\Delta_m(\theta) = \sum_{n=-\infty}^{\infty} \frac{\theta(n + m) - \theta(n)}{n}$$

(the prime signifies the omission of the term with $n = 0$; convergence of the series is assumed).

**Example 2.** Let us consider a non-empty compact set $S \subset \mathbb{C}$ whose plane Lebesgue measure is zero. The Hartogs-Rosenthal theorem on the denseness in

---

\(^{19}\)Further, any system of the form $(r^m)_{k=1}^{\infty}$, where $0 \leq n_1 < n_2 < \cdots$ and $\sum_{k=1}^{\infty} n_k^{-1} = \infty$, is complete in $C[0, 1]$ (theorem of M"untz).
C(S) of the set R(S) of rational functions with poles outside of S can be established
in the following way. Suppose that R(S) is annihilated by the linear functional
on C(S) defined by the complex measure μ which is concentrated on S. Then
$$\int_{c} \frac{d \mu(\zeta)}{z - \zeta} = \int_{s} \frac{d \mu(\zeta)}{z - \zeta} = 0 \quad (z \notin S).$$
We take any rectangle X for which μ(∂X) = 0 and λ₁(∂X ∩ S) = 0 (λ₁ is linear
Lebesgue measure). Then
$$\mu(X) = \frac{1}{2\pi i} \int_{c \setminus \partial X} d \mu(\zeta) \int_{\partial X} \frac{dz}{z - \zeta} = \frac{1}{2\pi i} \int_{\partial X \setminus S} dz \int_{c} \frac{d \mu(\zeta)}{z - \zeta} = 0.$$ 
Hence it follows that μ = 0.

A system of vectors (eᵢ)₀ for a separable LTS E is called a topological basis
or a Schauder basis if for each x ∈ E there is a unique expansion in a series:
$$x = \sum_{i=0}^{\infty} \xi_i e_i.$$ 
For example, in Lᴾ (1 ≤ p < ∞) the system (δᵢ)₀ is a Schauder basis. Topological
bases in C[0, 1] and Lᴾ(0, 1) (1 ≤ p < ∞) were described by Schauder in the
1920s. Later on much effort was expended on the construction of Schauder bases
in specific spaces²⁰ and also in the proof of the existence of such a basis in any
separable Banach space; however the last conjecture was disproved by Enflo
(1973). Soon after this counterexamples were constructed in subspaces of the
spaces $L_p$ (Davy (p > 2), Shankovskij (p < 2)).

Clearly, any Schauder basis is a complete linearly independent system²². The
coefficients $\xi_i$ in the expansion of a vector $x$ are called its coordinates (just as in
the algebraic situation). They are clearly linear functionals of $x$. The image $\hat{E}$ of
the homomorphism $x \mapsto (\xi_i(x))_{i=0}^{\infty}$ mapping $E$ into the space of sequences is
called the coordinate space of the given basis.

**Example.** Let $E$ be a separable Hilbert space and let $(e_i)₀$ be a complete
orthonormal system (it can be obtained by the orthogonalisation process from
any countable dense set $X \subseteq E$). This orthonormal basis is a Schauder basis in $E$.
The unique expansion of a vector is its Fourier series
$$x = \sum_{i=0}^{\infty} (x, e_i)e_i,$$
the coordinate space coincides with $l^2$ and an isometry between $E$ and $l^2$ is

²⁰ For example, the question of the existence of a Schauder basis in the space of functions which are
analytic on the unit disk and continuous up to the boundary remained open for a long time. A positive
solution was obtained by S.V. Bochkarev (1974).
²¹ It can be shown that in the class of Fréchet spaces a necessary condition for the existence of a
Schauder basis is that $E^*$ be total. Thus, for example, there is no Schauder basis in $L_P(0, 1)$ for
$0 < p < 1$.
²² It also has the stronger property of $\omega$-linear independence: $\sum_{i=0}^{\infty} \xi_i e_i = 0 \Rightarrow \xi_i = 0 \quad (i = 0, 1, 2, \ldots)$. 
established by means of Parseval's equality

$$\|x\|^2 = \sum_{i=0}^{\infty} |(x, e_i)|^2.$$ 

All this is proved in exactly the same way as for classical Fourier series in $L^2(-\pi, \pi)$. As a corollary we obtain the result that all separable Hilbert spaces are isometrically isomorphic to one another.

In the example considered the coordinate functionals are continuous and the situation is exactly the same for any Schauder basis $(e_i)^{\infty}_{\omega} \in$ a Banach space $E$ (see Section 3.8). Moreover the systems $(e_i)^{\infty}_{\omega} \in E$ and $(\xi_i)^{\infty}_{0} \in E^{\ast}$ are biorthogonal: $\xi_i(e_k) = \delta_{ik}$. Therefore the system $(e_i)^{\infty}_{\omega}$ is minimal, i.e. $e_k \notin \text{Lin}\{e_j; j \neq k\}$ ($k = 0, 1, 2, \ldots$). Conversely for any minimal system of vectors there exists a biorthogonal system of continuous linear functionals. We note that not every complete minimal system (even one with a total biorthogonal system) is a Schauder basis. In contrast to the Schauder basis, such a system exists in any separable Banach space (A.I. Markushevich, 1943).

A topological basis $(e_i)^{\infty}_{\omega}$ is said to be unconditional if all convergent series $\sum_{i=0}^{\infty} \xi_i e_i$ converge unconditionally, i.e. their convergence and sum do not depend on the order of the terms. For example, an orthonormal basis in a separable Hilbert space $H$ is such a basis (but there also exists in $H$ a conditional topological basis (K.I. Babenko, 1947)). There are no unconditional bases23 in $C[0, 1]$ (Karlin, 1948).

The system $(e^{i\lambda_n t})_{n \in \mathbb{Z}}$ in $L^2(-\pi, \pi)$ is an orthonormal basis for $\lambda_n = -n$ and it remains an unconditional topological basis for $\lambda_n$ sufficiently close to $n$.

**Theorem of M.I. Kadets.** If $(\lambda_n)_{n \in \mathbb{Z}}$ is a sequence of real numbers which satisfies the condition

$$\sup_n |\lambda_n - n| < \frac{1}{4},$$

then the corresponding system of exponentials $(e^{i\lambda_n t})_{n \in \mathbb{Z}}$ is an unconditional topological basis in $L^2(-\pi, \pi)$.

It is not possible to improve the constant $\frac{1}{4}$: if $\lambda_n = n - \frac{1}{4}\text{sgn} n$ the system even fails to be complete (Levinson, 1936).

On the subject of unconditional convergence of vector series $\sum_{i=0}^{\infty} x_i$ we note that in a Banach space this is equivalent to 1) convergence of all its 'parts' $\sum_{i=0}^{\infty} x_i e_i$, 2) convergence of all series $\sum_{i=0}^{\infty} e_i x_i (e_i = \pm 1)$. In a complete LCS $E$ a sufficient condition for unconditional convergence of a series $\sum_{i=0}^{\infty} x_i$ is that all series $\sum_{i=0}^{\infty} p(x_i)$ converge ($p$ runs through a family of basic seminorms). It is natural to call such convergence absolute. In a Banach space absolute convergence of a series means that $\sum_{i=0}^{\infty} ||x_i|| < \infty$. In any infinite-dimensional Banach space there exists a series which converges unconditionally but not absolutely

---

23 It is even possible to construct a reflexive separable Banach space having no unconditional basis (Pelczyński, 1969).
(Dvoretzki-Rogers, 1950) which cannot happen in the finite-dimensional case (theorem of Riemann). For a convergent series in a finite-dimensional space the set of sums obtained from all possible rearrangements which preserve convergence is an affine manifold (Steinitz, 1916). In any infinite-dimensional Banach space there is a counterexample to this (V.M. Kadets, 1986).

The concept of an unconditional topological basis can be generalised to the case of a non-separable LTS $E$ by applying this term to a system of vectors $(e_i)_{i \in I}$ with respect to which there exists for each $x \in E$ a unique unconditionally convergent series

$$x = \sum_{i \in I} \xi_i e_i$$

(the set of coordinates $\xi_i$ different from zero is no more than countable). Completeness and linear independence are clear in this case also. The coordinate space is now a subspace $\tilde{E} = \Phi(I, K)$ which is invariant with respect to rearrangements (it is symmetric) and moreover each $\xi \in \tilde{E}$ has no more than countable support.

**Example.** Let $E$ be any Hilbert space and let $B = (e_i)_{i \in I}$ be a complete orthonormal system (its existence can be shown with the help of Zorn’s lemma). This orthonormal basis is an unconditional topological basis. The unique expansion of any vector $x$ is its Fourier series

$$x = \sum_{i \in I} (x, e_i)e_i,$$

and the coordinate space coincides with $l^2 = L^2(I, \nu)$ ($\nu$ is the measure on $I$ such that $\nu(\{i\}) = 1$ for all $i \in I$); an isometry between $E$ and $l^2$ is established by means of Parseval’s equality

$$\|x\|^2 = \sum_{i \in I} |(x, e_i)|^2.$$

In particular, the system of exponentials $(e^{i\lambda t})_{\lambda \in \mathbb{R}}$ on the whole axis $-\infty < t < \infty$ is an orthonormal basis in the space $B^2$, the completion of the pre-Hilbert space $AP$ of almost-periodic functions. Thus $B^2$ is isometrically isomorphic to $l^2_\mathbb{R}$.

In any LTS $E$ the least power of all the complete systems is called its topological dimension and is denoted by $\dim_t E$. In the finite-dimensional case it is equal to its algebraic dimension, while in the infinite-dimensional case it is the least power of all the dense subsets. In particular, $E$ has countable topological dimension if and only if it is separable.

A system of vectors $(e_i)_{i \in I}$ in a linear metric space $E$ with metric $d$ is said to be $\delta$-distal if

$$d(e_i, e_k) \geq \delta > 0 \quad (i \neq k).$$

A system is said to be distal if it is $\delta$-distal for some $\delta$. The power of a distal system does not exceed $\dim_t E$ if $E$ is infinite-dimensional. In fact in each ball $d(x, e_i) < \delta/2$ there is a point $x_i$ from the dense subset and $x_i \neq x_k$ ($i \neq k$).

For an infinite-dimensional space $E$ the power of any complete distal system is equal to $\dim_t E$. 


In a Hilbert space $E$ any orthonormal system $(e_i)_{i \in I}$ is distal: $d(e_i, e_k) = \sqrt{2}$ ($i \neq k$); therefore the power of any orthonormal basis in $E$ is equal to $\dim_t E$. Consequently, all Hilbert spaces of the same topological dimension are isometrically isomorphic.

**Example.** The topological dimension of the space $B^2$ is equal to the power of the continuum.

Let $E$ be an infinite-dimensional normed linear space. By Zorn's lemma there exists for arbitrary $\delta (0 < \delta < 1)$ a maximal $\delta$-distal normalised system $(e_i)_{i \in I}$. It is complete since otherwise it would be possible by Riesz's lemma to extend it in such a way that the $\delta$-distal property is preserved. Therefore its power is equal to $\dim_t E$. This result remains valid if in place of $(e_i)_{i \in I}$ we construct by transfinite induction a system $(u_j)_{j \in J}$ with the following properties: 1) the index set $J$ is well-ordered; 2) $\|u_j\| = 1$ ($j \in J$); 3) $d(u_j, L_j) \geq \delta$ where $L_j$ is the closed linear hull of the subsystem $(u_k)_{k < j}$. If we now choose $f_j \in E^*$ such that $\|f_j\| = 1$, $|f_j(u_j)| > \delta/2$ and $f_j|_{L_j} = 0$, then the system $(f_j)_{j \in J}$ will be $\delta/2$-distal. Consequently, $\dim_t E^* \geq \dim_t E$.

If $E^*$ is separable then $E$ is also separable. If $E$ is reflexive then $\dim_t E = \dim_t E$.

The approach to the theory of topological dimension of normed spaces which has been described was proposed by M.G. Krejn, M.A. Krasnosel'skij and D.P. Milman (1948) as a basis for the investigation of those properties of subspaces which are stable under small perturbations.

It is natural to measure the size of a perturbation by means of the gap

$$\theta(L, M) = \max \left( \sup_{x \in S_L} d(x, M), \sup_{y \in S_M} d(y, L) \right),$$

where $S_L, S_M$ are the unit spheres of the subspaces $L, M$.

**Theorem 1.** If $\theta(L, M) = \theta < \frac{1}{2}$ then $\dim_t L = \dim_t M$.

**Proof.** Suppose that $(u_i)_{i \in I}$ is a maximal $\delta$-distal normalised system in $L$ with $2\theta < \delta < 1$. We take $\varepsilon < \delta/2 - \theta$ ($\varepsilon > 0$) and choose $v_i \in M$ such that $d(u_i, v_i) < \theta + \varepsilon$. Then the system $(v_i)_{i \in I}$ will be $[\delta - 2(\theta + \varepsilon)]$-distal. Consequently, $\dim_t L \leq \dim_t M$.

**Theorem 2.** If $\theta(L, M) = \theta < \frac{1}{2}$ and the subspaces $L, M$ are closed then $\dim_t(E/L) = \dim_t(E/M)$.

**Proof.** Suppose that $\dim_t(E/L) > \dim_t(E/M)$. We take in $E/L$ a maximal $\delta$-distal normalised system $(\{u_i\})_{i \in I}$, where $\{u_i\}$ is the class of the vector $u_i \in E$, $\|u_i\| < 1 + \varepsilon$ ($\varepsilon > 0$). Let us consider the class $\langle u_i \rangle$ modulo $M$. The system $(\{u_i\})_{i \in I}$ is not distal. Therefore for any $\eta > 0$ we can find $i_1, i_2 \in I$ such that $d(\langle u_{i_1} \rangle, \langle u_{i_2} \rangle) < \eta$ and then there exists $z \in M$ such that $d(u_{i_1} - u_{i_2}, z) < \eta$. Consequently

---

24 For the infinite-dimensional case. The finite-dimensional case requires a separate treatment with an application of topological methods.
\[ \theta(L, M) \geq d\left( \frac{z}{\|z\|}, L \right) \geq \frac{d(u_i, L) - d(u_i, z)}{\|z\|} \geq \frac{\delta - \eta}{2(1 + \varepsilon) + \eta}. \]

Letting \( \delta \to 1, \varepsilon \to 0, \eta \to 0 \) we obtain \( \theta(L, M) \geq \frac{1}{2} \), contrary to the given condition. \( \square \)

**Remark 1.** In the case of a Hilbert space it is sufficient in both theorems to require \( \theta(L, M) < 1 \).

**Remark 2.** The modified gap
\[
\tilde{\theta}(L, M) = \max \left( \sup_{x \in S_L} d(x, S_M), \sup_{y \in S_M} d(y, S_L) \right)
\]
is a complete metric on the set of subspaces of a Banach space (I. Ts. Gokhberg – A.S. Markus, 1959), whereas the original gap does not satisfy the triangle inequality. Moreover they are topologically equivalent because of the inequality
\[
1 \leq \tilde{\theta}(L, M)/\theta(L, M) \leq 2.
\]

**3.4. Extreme Points of Compact Convex Sets.** For any convex set \( X \) in a linear space \( E \) a point \( x \in X \) is called an extreme (or extremal) point if it does not belong to any interval \([x_1, x_2] \subset X \) whose end points are both different from \( x \). This is equivalent to the assertion that if \( x = (u + v)/2 \) where \( u, v \in X \) then \( u = v \). The set of extreme points of a convex set \( X \) is denoted by \( \text{extr} \ X \). Its role is contained in the following fundamental result.

**Krejn-Mil'man Theorem.** Let \( E \) be an LTS and suppose that \( E^\ast \) is total. Then any compact convex set \( Q \subset E \) coincides with the closed convex hull of the set \( \text{extr} \ Q \).

In this connection we note that the set of extreme points of a compact convex set need not be closed and hence it can fail to be compact.

**Example.** Let us consider in \( \mathbb{R}^3 \) any circle \( C \) and any interval \([x_1, x_2] \not\subset C \) whose end points are both different from \( x \). For the compact set \( Q = \text{Co}(C \cup [x_1, x_2]) \) the set of extreme points is \((C \setminus \{c\}) \cup \{x_1, x_2\}\).

Nevertheless, if \( E \) is finite-dimensional any compact convex set \( Q \subset E \) coincides with the convex hull of its set of extreme points. This assertion had already been proved by Minkowski (by induction on the dimension).

In any LTS the closure of a convex set is convex, from which it is clear that the closure of the convex hull of any set \( X \) is the smallest closed convex set containing \( X \). We call this the closed convex hull\(^{25}\) of the set \( X \). The convex hull of a set which is actually closed may fail to be closed.

**Example.** The convex hull of the set of points \((\xi, \xi^2) (\xi \geq 0) \) in \( \mathbb{R}^2 \) consists of zero and the points \((\xi, \eta) \) for which \( \eta \geq \xi^2 \) and \( \xi > 0 \). Its closure contains in addition the semiaxis \((0, \eta) (\eta > 0) \).

\(^{25}\)The closed absolutely convex hull is defined similarly.
§ 3. Linear Topology

In a finite-dimensional space the convex hull of any compact set is closed and consequently it is compact (its boundedness is obvious). In an infinite-dimensional space (even if it is a Hilbert space) the convex hull of a compact set may fail to be closed. In certain infinite-dimensional spaces there exist compact sets whose closed convex hulls are not compact. This is essentially the effect of incompleteness of such a space. In any complete LCS the closed convex hull of a compact set is compact. For a locally convex Fréchet space a similar assertion holds in the weak topology: the weak closure of the convex hull of a weakly compact set is weakly compact (Krejn-Shmul'yan theorem).

If the closed convex hull of a compact set \( \mathcal{Q} \) in an LTS \( E \) with total \( E^* \) is compact then all of its extreme points are contained in \( \mathcal{Q} \) (D.P. Mil'man, 1947).

We now give a proof of the Krejn-Mil'man theorem. First let us assume that \( E \) is an LCS (without loss of generality we can also assume that it is real). We put \( K = \text{Co}(\text{extr } \mathcal{Q}) \). If \( K \neq \mathcal{Q} \) then any point \( z \in \mathcal{Q} \setminus K \) can be separated from \( K \) by a closed hyperplane \( \{ x : f(x) = \alpha \} \) such that \( f(x) \leq \alpha (x \in K), f(z) > \alpha \). We put \( \beta = \max_{x \in \mathcal{Q}} f(x) \). Clearly \( \beta > \alpha \) and so the hyperplane \( H = \{ y : f(y) = \beta \} \) does not intersect \( K \) or, consequently, \( \text{extr } \mathcal{Q} \). This situation leads to a contradiction since we are able to find an extreme point of the compact set \( \mathcal{Q} \) in \( H \). For this purpose let us call a closed affine manifold \( A \) extreme if \( A \cap \mathcal{Q} \neq \emptyset \) and whenever an interior point of an interval \( [a, b] \subset \mathcal{Q} \) belongs to \( A \) then the entire interval is contained in \( A \). For example, the hyperplane \( H \) is extreme.

Using Zorn's lemma it is easy to show that there is a minimal extreme manifold \( M \) contained in \( H \). It is a singleton set since otherwise there exists a continuous linear functional \( g \) which is not constant on \( M \) and then the smaller extreme manifold \( \{ y : y \in M, g(y) = \gamma \} \), where \( \gamma = \max_{x \in M \cap \mathcal{Q}} g(x) \), is contained in \( M \). But any singleton extreme manifold must simply consist of an extreme point of the compact set \( \mathcal{Q} \).

Now if we only know that \( E^* \) is total, from what has been proved the theorem is true for the weak topology, i.e. \( \mathcal{Q} = \hat{\mathcal{C}} \) where \( \hat{\mathcal{C}} \) is the weak closure of the set \( \mathcal{C} = \text{Co}(\text{extr } \mathcal{Q}) \). Denoting the strong closure by \( \mathcal{C} \) we have \( C \subset \mathcal{C} \subset \hat{\mathcal{C}} \). But \( C \) is compact, consequently weakly compact and so also weakly closed. Therefore \( \overline{C} = \hat{\mathcal{C}} \), i.e. \( \mathcal{Q} = \hat{\mathcal{C}} \). \( \square \)

It is not possible to dispense with the condition that \( E^* \) be total: there exists an LTS in which a certain non-empty compact convex set has no extreme point (Roberts, 1977).

The condition of compactness in the Krejn-Mil'man theorem is also essential; for example, it is not possible to replace it by closedness and boundedness.

**Example.** If \( S \) is a connected compact space the closed unit ball \( \| \varphi \| \leq 1 \) in the Banach space of real continuous functions on \( S \) has no extreme points other than the constants \( \pm 1 \).

\(^{26}\)It is also the closure in the strong (original) topology since the space is locally convex.
In this connection we stress that the closed unit ball of an infinite-dimensional Banach space is not compact, since otherwise the space would be locally compact and consequently finite-dimensional.

One of the many useful corollaries of the Krejn-Mil'man theorem is that under its conditions the maximum of any continuous convex functional on a compact convex set is attained on its set of extreme points.

3.5. Integration of Vector-Functions and Measures. The initial development of this important analytic tool was carried out in the 1930s. We give the Gel'fand-Pettis definition of the integral of a vector-function (1938). Let $S$ be a set with a measure $\mu$, $E$ an LTS with total $E^*$ and $x(s) (s \in S)$ a vector-function, i.e. a function with values in $E$. Let $x(s)$ be integrable in the sense that all scalar functions of the form $f(x(s)) (f \in E^*)$ are summable with respect to the measure $\mu$. Then the linear functional

$$J(f) = \int_S f(x(s)) \, d\mu$$

is called the integral of $x(s)$ with respect to the measure $\mu$. Thus, in general, $J \in (E^*)^*$, however the most interesting case is when $J(f) = f(\bar{x})$ where $\bar{x} \in E$. If such an $\bar{x}$ exists it is unique; it can be identified with $J$ and we put correspondingly

$$\bar{x} = \int_S x(s) \, d\mu.$$

**Theorem.** Suppose that $\mu(S) < \infty$ and $x(s)$ is an integrable vector-function. If the closed convex hull $X$ of the set of values $x(S)$ is compact then the integral of $x(s)$ with respect to the measure $\mu$ is a vector $\bar{x} \in E$ (further, $\bar{x} \in X$ if $\mu(S) = 1$).

**Proof.** We may assume that $\mu(S) = 1$ and that $E$ is a real space. For any collection $\{f_i\}_1^n \subset E^*$ we consider the continuous mapping $F x = (f_i(x))_1^n$ from $X$ into $\mathbb{R}^n$. Its image is a compact convex set. Moreover $(J(f_i))_1^n \in \text{Im} F$ since otherwise there exists a vector $(\alpha_i)_1^n$ such that

$$\sum_{i=1}^n \alpha_i J(f_i) > \sup_{x \in X} \sum_{i=1}^n \alpha_i f_i(x),$$

i.e. for the functional $g = \sum_{i=1}^n \alpha_i f_i$ the inequality

$$\int_S g(x(s)) \, d\mu > \sup_{x \in X} g(x) \geq \sup_{x \in S} g(x(s))$$

is satisfied, which is absurd. If we now consider the closed hyperplanes $H_f = \{z : f(z) = J(f)\} (f \in E^*)$ in $E$, we see that the family of compact sets $H_f \cap X$ has the finite intersection property and so its intersection $X_0$ is non-empty. The vector $\bar{x} \in X_0$ is what is required. □

Integrability with respect to a generalised measure $\omega$ is defined as integrability with respect to each of its components and, correspondingly, the preceding
§ 3. Linear Topology

Theorem can be formulated componentwise. In particular it is true if the
generalised measure $\omega$ is absolutely continuous with respect to some measure $\mu$
(infinite in general),

$$\omega(A) = \int_A \rho(s) \, d\mu \quad (\rho \in L^1(S, \mu), A \subset S),$$

while the function $x(s)$ satisfies the previous conditions.

In the most developed situation we have to integrate a function $x(s)$ with values
in a Banach space $E$ which is defined and weakly continuous on a locally compact
topological space $S$ provided with a generalised measure $\omega$. In this situation the
norm of the function is measurable since $\|x(s)\| = \sup |f(x(s))| \ (f \in E^*, \|f\| = 1)$
and all the $f(x(s))$ are continuous. If

$$v(x) = \int \|x(s)\| \, |d\omega| < \infty$$

then $x(s)$ is integrable (in this case it is natural to say that it is absolutely
integrable and the integral

$$\bar{x} = \int_S x(s) \, d\omega \in E^{**}$$

is a vector in $E$; moreover $\|\bar{x}\| \leq v(x)$. In fact for any $\varepsilon > 0$ we can choose a
compact set $Q_\varepsilon \subset S$ such that

$$\int_{S \setminus Q_\varepsilon} \|x(s)\| \, |d\omega| < \varepsilon.$$ 

The set $x(Q_\varepsilon)$ is weakly compact. By the Krejn-Shmul'yan theorem its weakly
closed convex hull is also weakly compact. Consequently

$$\bar{x}_\varepsilon = \int_{Q_\varepsilon} x(s) \, d\omega \in E.$$ 

Further

$$\|\bar{x} - \bar{x}_\varepsilon\|_{E^{**}} = \sup_{f \in E^*, \|f\| = 1} \left| \int_{S \setminus Q_\varepsilon} f(x(s)) \, d\omega \right| < \varepsilon.$$ 

Since $E$ is closed in $E^{**}$ we deduce that $\bar{x} \in E$.

Let $S$ be any space with a measure $\mu$. Following Bochner (1933) we can define
the Banach space $L^1(S, \mu; E)$ of absolutely integrable (summable) vector-functions
with values in the Banach space $E$ as the $L^1$-completion of the space of simple
functions, i.e. finite combinations of the form $\sum_k a_k \chi_k(s)$, where the $a_k$ are arbitrary vectors in $E$ and the $\chi_k$ are characteristic functions of measurable sets $S_k$ such that $\mu(S_k) < \infty$ and $S_i \cap S_k = \emptyset$ ($i \neq k$). Moreover if $x \in L^1(S, \mu; E)$ the function $\|x(s)\|$ is measurable and
\[ \|x\| = \int_S \|x(s)\| \, d\mu. \]

On $L^1(S, \mu; E)$ the Bochner integral of $x$ with respect to the measure $\mu$ is defined as the strong limit of integrals of simple functions and
\[ \int_S \left( \sum_k a_k \chi_k(s) \right) \, d\mu = \sum_k a_k \mu(S_k) \in E. \]

The Lebesgue theory of integration extends to this situation with no significant changes. All functions in $L^1(S, \mu; E)$ are Gelfand-Pettis integrable and the values of the integrals in both senses coincide in this case.

If $x \in L^1(S, \mu; E)$ the set function
\[ m_x(A) = \int_A x(s) \, d\mu \quad (A \subset S) \]
on the $\sigma$-algebra associated with the measure $\mu$ is countably additive and in this sense it is a vector measure. For any vector measure $m(A)$ its variation is the measure
\[ |m|(A) = \sup_{n} \sum_{k=1}^{n} \|m(A_k)\| \leq \infty, \]
where $\{A_k\}^n$ runs through all finite measurable partitions of the set $A$. In the original example
\[ |m_x|(A) = \int_A \|x(s)\| \, d\mu. \]

The total variation (or norm) of a vector measure $m$ is the quantity $\|m\| = |m|(S)$. A vector measure $m$ is said to be absolutely continuous if $\|m\| < \infty$ and
\[ m(A) = \int_A x(s) \, d|m|, \]

where $x \in L^1(S, |m|; E)$. We say that a real Banach space $E$ has the Radon-Nikodým property (RN) if every vector measure with finite norm is absolutely continuous\textsuperscript{30}. The RN property has a purely geometric characterisation which is closely connected with the Krejn-Mil'man theorem.

\textsuperscript{29} Any summable function is the limit almost everywhere of some sequence of simple functions. Therefore its set of values is separable (except possibly for values taken on a set of measure zero).

\textsuperscript{30} The classical Radon-Nikodým theorem asserts that if there are two measures $\mu, \nu$ (on a common $\sigma$-algebra) and $\mu(A) = 0 \Rightarrow \nu(A) = 0$ then
\[ \nu(A) = \int_A \rho \, d\mu, \]
where $\rho$ is some function (the density or Radon-Nikodým derivative).
Let \( M \subset E \) be a closed convex bounded set. A point \( z_0 \in M \) is said to be **strongly exposed** if there exists a functional \( f \in E^* \) such that \( f(z) < f(z_0) \) for all \( z \in M \ (z \neq z_0) \) and the diameter of the set of \( \{z \in M \mid f(z) > f(z_0) - \varepsilon \} \) tends to zero as \( \varepsilon \to 0 \).

**Phelps-Rieffel Theorem.** A necessary and sufficient condition for a Banach space to have property \( \text{RN} \) is that each of its closed bounded convex subsets be the closed convex hull of its set of strongly exposed points.

Since all strongly exposed points are clearly extreme\(^{31}\), in an \( \text{RN} \)-space each closed bounded convex set is the closed convex hull of its set of extreme points\(^{32}\) (Lindenstrauss, 1966). The converse assertion holds if \( E \) is a conjugate space (Huff-Morris, 1976). Each separable conjugate Banach space has the property \( \text{RN} \) (Dunford-Pettis, 1940).

We now use the tools of integration to give another formulation of the Krejn-Mil'man theorem. Let \( E \) be an LTS with total \( E^* \), let \( Q \subset E \) be a compact convex set and put \( S = \text{extr } Q \). Let us consider on the compact set \( S \) a normalised measure \( \mu \) and the function \( x(s) = s \). The *centre of mass* \( x_\mu = \int_S s \, d\mu \) belongs to \( Q \) and the Krejn-Mil'man theorem implies that any point \( \bar{x} \in Q \) is the centre of mass of some normalised measure \( \mu_{\bar{x}} \) on \( S \). A significant refinement of this result is due to Choquet (1956): if the compact set \( Q \) is metrizable then \( \mu_{\bar{x}} \) (\( \bar{x} \in Q \)) can be chosen to be concentrated\(^{33}\) on \( \text{extr } Q \) and not on its closure \( S \) (in this situation \( \text{extr } Q \) is a \( G_\delta \) set). Such a \( \mu_{\bar{x}} \) is called a **Choquet measure** for the point \( \bar{x} \). It can be regarded as a solution of a singular abstract moment problem. This way of looking at it embraces in a single form all those classical moment problems in which the required measure is finite (the general case needs an additional limiting process). For example, in the power moment problem on the whole axis we have a convex set \( Q \) of positive (in the sense of the corresponding Hankel matrix) sequences \( (m_k)_{k=0}^{\infty} \) \((m_0 = 1)\), its extreme points are \( \{t^k\}_{k=0}^{\infty} \) \((t \) is a real parameter) and the integral representation

\[
m_k^\text{H} = \int_0^\infty t^k \, d\sigma(t) \quad (k = 0, 1, 2, \ldots)
\]

corresponds to Choquet's theorem.

---

\(^{31}\) The converse is false. However, as we see, strongly exposed points play the same role as extreme points in a series of cases. We note here that any weakly compact convex set in a Banach space is the closed convex hull of its set of strongly exposed points (Amir-Lindenstrauss, 1968). In a uniformly convex Banach space every point of the unit sphere is a strongly exposed point of the closed unit ball.

\(^{32}\) Thus, for example, if \( S \) is a connected compact space with \( \text{card } S > 1 \) then \( C(S) \) is not an \( \text{RN} \)-space.

\(^{33}\) We say that a measure \( \nu \) on a set \( X \) is *concentrated* on a subset \( M \subset X \) if \( \nu(X \setminus M) = 0 \). If \( X \) is a topological space the smallest closed subset on which the measure \( \nu \) is concentrated is called its *support* and is denoted by \( \text{supp } \nu \).
The integration of vector-functions also plays a significant role in the extension of complex analysis to the vector situation, which in turn is absolutely essential for constructing spectral theory.

A vector-function \( x(\zeta) \) defined on some domain \( G \) of the complex plane \( \mathbb{C} \) and taking values in an LTS \( E \) with total \( E^* \) is said to be analytic on \( G \) if all scalar functions of the form \( f(x(\zeta)) \) (\( f \in E^* \)) are analytic on \( G \).\(^{34}\)

**Theorem.** If \( x(\zeta) \) is an analytic function on a domain \( G \) with values in a Banach space \( E \) then at each point \( \zeta_0 \in G \) it has a strong expansion as a power series

\[
x(\zeta) = \sum_{k=0}^{\infty} x_k(\zeta - \zeta_0)^k \quad (|\zeta - \zeta_0| < \delta = d(\zeta_0, \mathbb{C} \setminus G)),
\]

in which the vector coefficients are uniquely determined.

**Proof.** We have

\[
f(x(\zeta)) = \sum_{k=0}^{\infty} \xi_k(f)(\zeta - \zeta_0)^k
\]

(\( f \in E^* \)),

where

\[
\xi_k(f) = \frac{1}{2\pi i} \int_{|\zeta - \zeta_0|=\rho} \frac{f(x(\zeta))}{(\zeta - \zeta_0)^{k+1}} \quad (0 < \rho < \delta).
\]

Since \( x(\zeta) \) is weakly continuous on the circle \( |\zeta - \zeta_0| = \rho \), we have \( \xi_k(f) = f(x_k) \), where \( x_k \in E \) (\( k = 0, 1, 2, \ldots \)).

Moreover \( \|x_k\| \leq M(\rho)\rho^{-k} \), where \( M(\rho) = \sup_{|\zeta - \zeta_0|=\rho} \|x(\zeta)\| \). Consequently, the power series (17) converges absolutely on the disk \( |\zeta - \zeta_0| < \rho \) for any \( \rho < \delta \), i.e. finally, on the disk \( |\zeta - \zeta_0| < \delta \). Its sum \( s(\zeta) \equiv x(\zeta) \) since \( f(s(\zeta)) = f(x(\zeta)) \) for all \( f \in E^* \).

The further development of the theory of analytic vector-functions with values in a Banach space presents no problem. Concerning this we note only the Cauchy-Hadamard formula for the radius of convergence of a power series with vector coefficients \( c_k (k = 0, 1, 2, \ldots) \):

\[
r = \left( \lim_{k \to \infty} \frac{k}{\|c_k\|} \right)^{-1}.
\]

The sum of the power series is analytic inside the disk of radius \( r \).

### 3.6. \( w^* \)-Topologies

Let us consider an arbitrary linear space \( E \) and its algebraic conjugate space \( E^* \). Since \( E^* \subset \Phi(E, K) \) the topology of pointwise convergence is defined on \( E^* \). It is called the \( w^* \)-topology on \( E^* \). The LCS \( E^* \) is \( w^* \)-closed in \( \Phi(E, K) \) and consequently it is \( w^* \)-complete.

If \( E \) is an LTS, its conjugate space \( E^* \) is contained in \( E^\sharp \) and so the \( w^* \)-topology is also defined on \( E^\sharp \). This is the weakest topology in which all the linear

\(^{34}\) A vector-function which is analytic on the whole plane is said to be entire.
functionals of the form $\mathcal{A}(f) = f(x)$ ($x \in E, f \in E^*$) are continuous (these are the canonical images of the elements $x \in E$ in the space $(E^*)^\#$). Each $w^*$-continuous linear functional $E^*$ has the stated form.

A fundamental result of a general character concerning the $w^*$-topology is the

**Banach-Alaoglu Theorem.** Let $E$ be an LTS and $U \subseteq E$ a neighbourhood of zero. Then the set $U^0$ consisting of those linear functionals on $E$ which satisfy the condition $\sup_{x \in U} |f(x)| \leq 1$ (and so are continuous) is $w^*$-compact.

**Proof.** We put $\Delta = \{ \lambda : |\lambda| \leq 1 \}$ and consider the set of functions $\Phi(U, \Delta)$ with its product topology. This is compact by Tikhonov’s theorem. The mapping $Rf = f|_U$ of the set $U^0 \subseteq E^*$ provided with the $w^*$-topology into $\Phi(U, \Delta)$ is a homeomorphism between $U^0$ and $RU^0$. But, as is easily verified, $RU^0$ is closed. Consequently it is compact. □

From the Banach-Alaoglu theorem and the Krejn-Mil’man theorem follows

**Corollary 1.** $U^0$ coincides with the $w^*$-closed convex hull of the set $\text{extr} U^0$.

Now let $E$ be a normed space. Then we have on $E^*$ at least three natural locally convex topologies: the strong topology defined by the norm of the linear functionals (with respect to which $E^*$ is a Banach space); the weak topology defined by the family of strongly continuous, i.e. bounded (for the norm), linear functionals; and the $w^*$-topology. Applying the Banach-Alaoglu theorem to the open unit ball $D \subseteq E$ we obtain the closed unit ball of $E^*$ as $D^0$. Thus we have established

**Corollary 2.** If $E$ is a normed space the closed unit ball in the Banach space $E^*$ is $w^*$-compact.

By the Krejn-Mil’man theorem the ball $D^0$ is the $w^*$-closed convex hull of its set of extreme points. Hence it is clear, incidentally, that a Banach space need not be the conjugate of some normed space; for example the spaces $C[0, 1], L^1[0, 1]$ are not conjugate spaces (even up to isomorphism).

The $w^*$-compactness of the closed unit ball in the space of measures $C(S)^*$ ($S$ compact) is the content of a theorem of Helly.

The spaces of continuous functions on compact sets have the special property of universality with respect to the class of all Banach spaces. This also follows from the Banach-Alaoglu theorem.

**Corollary 3.** Any Banach space $E$ is isometrically isomorphic to a closed subspace of the space of continuous functions on some compact set.

---

35 For any non-empty $X \subseteq E$ the set $X^0 = \{ f : f \in E^*, \sup_{x \in X} |f(x)| \leq 1 \}$ is called the polar of the set $X$. The polars of a set and of its absolutely convex hull clearly coincide. It is also clear that $X^0$ is absolutely convex and $w^*$-closed. The term ‘polar’ is also used in a relative sense, i.e. for the intersection of $X^0$ with subspaces of $E^*$ (in particular, with $E^*$). The notation $X^0$ is retained in this situation. If $X$ is a subspace then $X^0$ is its annihilator: $X^0 = X^\perp$.

36 This is also often called weak in view of the fact that it is weaker than the norm topology. In order to avoid confusion we will not do this (sometimes we call the $w^*$-topology the weakened topology).
Proof. The canonical isometry $\nu: E \to E^{**}$ associates with each $x \in E$ the function $(\nu x)(f) = f(x)$ which is $w^*$-continuous on the ball $D^0$. The norm of this function in $C(D^0)$, i.e. $\max_{f \in D^0} |f(x)|$, is equal to $\|x\|$. □

If $E$ is separable the $w^*$-topology on the unit ball $D^0 \subset E^*$ is metrizable. Now, as is well-known, any compact metric space is the image of the Cantor set $\mathcal{G} \subset [0, 1]$ under some continuous mapping. Thus in the case under consideration $C(D^0)$ is isometrically isomorphic to a subspace of $C(\mathcal{G})$ and this in turn is isometrically isomorphic to a subspace of $C[0, 1]$. Thus we have the

**Banach-Mazur Theorem.** Any separable Banach space is isometrically isometric to a closed subspace of $C[0, 1]$.

We give a further application of the metrizability of the $w^*$-topology on $D^0$ in the separable case. Let $E$ be a separable Banach space with a closed cone $K$. Then there exists a continuous linear functional $f > 0$ (M.G. Krejn, M.A. Rutman, 1948). For the proof we note that the compact set $D^0$, being metrizable, is separable. Consequently there is also a countable dense subset $\{f_n\}_{n=1}^{\infty}$ of the intersection $D^0 \cap K^*$. It just remains to put, for example, $f = \sum_n 2^{-n} f_n$.

**Remark.** Separability of the space $E$ in the last assertion is essential.

Any Banach space $E$ coincides with the subspace of $E^{**}$ consisting of all $w^*$-continuous linear functionals on $E^*$. Thus if $E$ is reflexive (and only in this case) each continuous linear functional on $E^*$ is $w^*$-continuous.

The weak topology on $E$ is induced by the $w^*$-topology of the space $E^{**}$. With respect to this topology $E$ is dense in $E^{**}$, since for any $\xi \in E^{**}$ and any finite collection $(f_i)^n_1 \subset E^*$ the system of linear equations $f_i(x) = \xi(f_i)$ ($1 \leq i \leq n$) has a solution $x \in E$. The unit ball $D \subset E$ is also $w^*$-dense in the unit ball $D^{**} \subset E^{**}$ which contains it, since otherwise there exists an element $\xi \in D^{**}$ which is separated from $D$ by a $w^*$-continuous linear functional; this is a functional $f \in E^*$ such that $\xi(f) \neq 0$ and $f|_D = 0$ — a contradiction.

**Theorem.** For reflexivity of a Banach space $E$ it is necessary and sufficient that the ball $\|x\| \leq 1$ be weakly compact.

**Necessity.** This follows from Corollary 2 of the Banach-Alaoglu theorem.

**Sufficiency.** The image of the closed ball in $E^{**}$ is $w^*$-compact and consequently $w^*$-closed. On the other hand it is $w^*$-dense in the closed ball of the space $E^{**}$. These balls therefore coincide and then we also have $E = E^{**}$. □

**Corollary.** A necessary and sufficient condition for each closed bounded convex set in a Banach space $E$ to be weakly compact is that $E$ be reflexive.

This result is extremely useful in the theory of extremal problems, in particular in the theory of approximation. Reverting to the latter, let us consider a non-empty closed subset $Y$ in any metric space $E$. For any point $x_0$ an element $y_0 \in Y$ is called a best approximation in $Y$ if
If $E$ is a reflexive Banach space then for any non-empty closed convex set $Y \subset E$ and any point $x_0 \in E$ a best approximation $y_0 \in Y$ exists. In fact it is sufficient to look for $y_0$ in the intersection $Y_0$ of the set $Y$ with the ball $d(x_0, y) \leq d(x_0, Y) + 1$. Now $Y_0$ is weakly compact and the functional $d(x_0, y) - \|x_0 - y\|$ is weakly lower semicontinuous. Therefore $d(x_0, y)$ attains its infimum on $Y_0$. The condition of reflexivity is necessary because of James criterion. We note in passing that the theorem on the uniqueness of the best approximation for all $x_0 \in E$ and all convex $Y \subset E$ holds if and only if $E$ is strictly convex. Thus in a uniformly convex Banach space $E$ any non-empty closed convex set $Y \subset E$ (in particular, any closed subspace) is a Chebyshev set in the sense that a best approximation $y_0 \in Y$ exists and is unique for each point $x_0 \in E$. N.V. Efimov and S.B. Stechkin have investigated the general problem of the convexity of Chebyshev sets (1961).

**Example.** Let us consider the problem of the best approximation in a Hilbert space $E^{38}$. Let $L \subset E$ be a closed subspace, let $x_0 \in E$ and let $y_0$ be the best approximation in $L$ for $x_0$; put $z_0 = x_0 - y_0$. Then for any $y \in L$ the function

$$\varphi(\tau) = \|z_0 + \tau y\|^2 = \|z_0\|^2 + 2\tau \text{re}(z_0, y) + \tau^2 \|y\|^2 \quad (\tau \in \mathbb{R})$$

has a minimum at the point $\tau = 0$, hence $\text{re}(z_0, y) = 0$. Consequently, $(z_0, y) = 0$ (in the real case this is immediate, while in the complex case it comes from substituting $iy$ for $y$ in the result). Thus $y_0 \in L$ and $x_0 - y_0$ is orthogonal to the subspace $L$ (i.e. to every vector in $L$). These properties characterise $y_0$ geometrically as the orthogonal projection of the vector $x_0$ on the subspace $L$. The value $d_0$ of the deviation of $x_0$ from $L$ is calculated by Pythagoras's theorem: $d_0 = \sqrt{\|x_0\|^2 - \|y_0\|^2}$.

### 3.7. Theory of Duality

Let $E$ and $F$ be two linear spaces and let $\langle x, y \rangle (x \in E, y \in F)$ be a bilinear functional such that

$$\forall x \neq 0 \exists y \text{ with } \langle x, y \rangle \neq 0 \text{ and } \forall y \neq 0 \exists x \text{ with } \langle x, y \rangle \neq 0.$$ 

Then we say that there is a duality between $E$ and $F$ or that $(E, F)$ is a dual pair of spaces.

**Example 1.** Let $E$ be any linear space, let $F = E^*$, its algebraic conjugate, and put $\langle x, f \rangle = f(x)$. Clearly this is a dual pair.

**Example 2.** If $E$ is an LTS and $E^*$ is total (for example, if $E$ is an LCS) then $(E, E^*)$ is a dual pair with $\langle x, f \rangle = f(x)$.

If $(E, F)$ is a dual pair there is a natural embedding of $F$ in $E^*$ under which $F$ is total on $E$. Thus $F$ defines a locally convex topology on $E$ which is denoted

---

37 $d(x_0, y_0)$ is called the *deviation* of the point $x_0$ from the set $Y$.

38 i.e. we are concerned with the *method of least squares*. 
by \( \sigma(E, F) \). The space \( E^* \), the conjugate of \( E \) in this topology, coincides with \( F \).

Now we can consider the \( w^* \)-topology on \( F = E^* \). Clearly it coincides with the topology \( \sigma(F, E) \) for the dual pair \((F, E)\) and consequently \( F^* = E \). In this way a complete symmetry between the linear topology spaces\(^{39}\) \( E \) and \( E^* \) is obtained. This symmetry also appears in the following property of reciprocity of polars.

**Theorem.** Let \((E, F)\) be a dual pair, \( X \subset E \) \((X \neq \emptyset)\) and \( Y = X^0 \), the polar of \( X \) in \( F \). Then \( Y^0 \), the polar of \( Y \) in \( E \), coincides with the \( \sigma(E, F) \)-closed absolutely convex hull \( \tilde{X} \) of the set \( X \). In particular, if \( X \) is absolutely convex and \( \sigma(E, F) \)-closed then \( X = Y^0 \) (i.e. \( X^{00} = X \)).

**Proof.** Since \( \langle x, y \rangle \leq 1 \) \((x \in X, y \in Y)\) we have \( X \subset Y^0 \) and hence \( \tilde{X} \subset Y^0 \).

Let \( x_0 \notin \tilde{X} \). By the theorem on the separating hyperplane, there exists \( y \in F \) such that \( \langle x, y \rangle \leq 1 \) \((x \in \tilde{X})\) and \( \langle x_0, y \rangle > 1 \). But then \( y \in Y \) and consequently \( x_0 \notin Y^0 \). It follows that \( \tilde{X} = Y^0 \). \( \square \)

**Corollary 1.** Let \((E, F)\) be a dual pair, \( L \) a subspace of \( E \) and \( N = L^\perp \), the annihilator of \( L \) in \( F \). Then the annihilator \( N^\perp \) of \( N \) in \( E \) coincides with the \( \sigma(E, F) \)-closure of \( L \). In particular, if \( L \) is \( \sigma(E, F) \)-closed then \( L = N^\perp \), i.e. \((L^\perp)^\perp = L \).

We note further that, clearly, for any family of sets \( \{X_i\} \) in \( E \)

\[
\left( \bigcup_i X_i \right)^0 = \bigcap_i X_i^0 .
\]

Hence on replacing \( X_i \) by \( X_i^0 \) we obtain from the previous theorem

**Corollary 2.** For any family of \( \sigma(E, F) \)-closed absolutely convex sets \( X_i \subset E \) the polar of their intersection coincides with the \( \sigma(F, E) \)-closed absolutely convex hull of the union of their polars \( X_i^0 \).

If \((E, F)\) is a dual pair we say about a locally convex topology on \( E \) for which \( E^* = F \) that it is consistent with the given duality. The weakest among such topologies is \( \sigma(E, F) \). It turns out that the class of topologies which are consistent with the given duality can be described and, moreover, among these topologies there is a finest.

**Mackey-Arens Theorem.** Let \((E, F)\) be a dual pair. In order that a locally convex topology on \( E \) be consistent with the given duality it is necessary and sufficient that it be a topology of uniform convergence on some family of absolutely convex subsets of the space \( F \) which are compact in the topology \( \sigma(F, E) \).

**Necessity.** The polar \( U^0 \) of any neighbourhood of zero \( U \subset E \) is compact in the topology \( \sigma(F, E) \) by the Banach-Alaoglu theorem, since by assumption \( F = E^* \) and \( \sigma(F, E) \) is the \( w^* \)-topology. The initial topology \( \tau \) on \( E \) coincides

\(^{39}\)In this connection, if \( E \) is an LTS to start with, its initial topology is to be replaced by the weak topology; moreover \( E^* \) is preserved and it is provided with the \( w^* \)-topology.
with the topology of uniform convergence on the family \( \{U^0\} \) where \( U \) runs through the family \( T \) of absolutely convex (open) \( \tau \)-neighbourhoods of zero in \( E \). In fact, if \( U \in T \) and

\[
V_U = \left\{ x: x \in E, \sup_{y \in U^0} |\langle x, y \rangle| < 1 \right\},
\]

then \( U \subseteq V_U \), since \( |\langle x, y \rangle| < 1 \) (\( x \in U \), \( y \in U^0 \)), \( U^0 \) is compact and the function \( \varphi_x(y) = |\langle x, y \rangle| \) on \( U^0 \) is continuous and, consequently, attains its maximum.

On the other hand if \( W \in T \) we can find \( U \in T \) such that its \( \sigma(E, E^*) \)-closure \( \overline{U} \) is contained in \( W \). Thus \( \overline{U} = U^{oo} \supseteq V_U \) and so \( V_U \subseteq W \).

**Sufficiency.** The topology \( \rho \) being considered on \( E \) is stronger than \( \sigma(E, F) \), hence \( F \subseteq E^* \). If \( f \in E^* \) then \( |f(x)| < 1 \) on a \( \rho \)-neighbourhood

\[
U_Q = \{ x: x \in E, |\langle x, y \rangle| < 1 \ (y \in Q) \}
\]

\((Q \) is an absolutely convex \( \sigma(F, E) \)-compact subset of \( F \)). Consequently \( |f(x)| \leq 1 \) for \( x \in Q^0 \), i.e. \( f \in Q^{oo} \) (the second polar is taken in \( E^* \)). Since \( Q \) is also compact in the space \( E^* \supseteq F \) and therefore closed in \( E^* \) we have \( Q^{oo} = Q \). Thus \( f \in Q \) and hence \( f \in F \). This shows that \( E^* = F \).

**Corollary.** If \( (E, F) \) is a dual pair the strongest topology on \( E \) which is consistent with the given duality is the topology of uniform convergence on all absolutely convex \( \sigma(F, E) \)-compact subsets of \( F \).

This topology is called the **Mackey topology** and is denoted by \( \tau(E, F) \). An LCS \( E \) whose topology coincides with \( \tau(E, E^*) \) is called a **Mackey space**. Any barrelled (and also any metrizable) LCS is a Mackey space.

**Remark.** If \( (E, F) \) is a dual pair, the closure of any convex set \( X \subseteq E \) is the same for all topologies which are consistent with the given duality since, as a result of the theorem on the separating hyperplane, the operation of closure on the class of convex sets is determined by the set of continuous linear functionals. Taking this as his starting point, D.A. Rajkov (1962) constructed a purely algebraic theory of duality and obtained an algebraic variant of the Krejn-Milman theorem.

Many important questions in the theory of duality for LTSs are connected with the consideration of bounded sets.

**Theorem of Mackey.** If \( (E, F) \) is a dual pair the same sets are bounded for all topologies on \( E \) which are consistent with the given duality.

**Proof.** By the Mackey-Arens theorem it is sufficient to show that each \( \sigma(E, F) \)-bounded set \( X \subseteq E \) is \( \tau(E, F) \)-bounded. Moreover we need consider only the case where \( X \) is countable: \( X = \{x_n\}_{n \in \mathbb{N}} \). If it is not \( \tau(E, F) \)-bounded there exists \( \{a_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}^+ \) such that \( \lim_{n \to \infty} a_n = 0 \) but \( \{a_n x_n\}_{n \in \mathbb{N}} \) does not tend to zero in the topology \( \tau(E, F) \). Then we can find \( \varepsilon > 0 \) and an absolutely convex \( \sigma(F, E) \)-compact set \( Q \subseteq F \) such that \( \max_{y \in Q} |\langle a_n x_n, y \rangle| \geq \varepsilon \) for an infinite set of suffices \( n \) (we can suppose that this holds for all \( n \)). There is therefore a sequence
\{y_n\}_{n=1}^{\infty} \subset Q \text{ such that } \langle x_n, y_n \rangle \to \infty. \text{ At the same time, sup}_n |\langle x_n, y \rangle| < \infty \text{ for each } y \in F \text{ because } X \text{ is } \sigma(E, F)\text{-bounded. This situation contradicts the Banach-Steinhaus theorem.} \square

Corollary. In any LCS weak boundedness of a set is equivalent to its boundedness.

We introduce for any LTS $E$ the uniform topology $\beta(E^*, E)$ on $E^*$ which is defined to be the topology of uniform convergence on all bounded subsets $X \subset E$. For a normed space $E$ this topology coincides with the norm topology of $E^*$, i.e. its strong topology. Clearly $\beta(E^*, E)$ is always finer than the $w^*$-topology40 and the strengthening of the topology can lead to an enlargement of the conjugate space.

The space of uniformly continuous linear functionals on $E^*$ is called the second conjugate of $E$ and is denoted by $E^{**}$. The standard canonical homomorphism $\nu: E \to E^{**}$ is injective if $E^*$ is total. In this case we have an embedding $E \subset E^{**}$ in place of the equality under the weak topology. The space $E$ is said to be (topologically) reflexive if $E = E^{**}$ and the initial topology on $E$ coincides with $\beta(E^{**}, E^*)$. In other words, reflexivity of $E$ means that $\nu$ is an isomorphism of the LTSs $E$ and $E^{**}$. Moreover $E$ is necessarily an LCS. If $E$ is reflexive, it is clear that $E^*$ is also reflexive under its uniform topology.

Theorem. In order that an LCS $E$ be reflexive it is necessary and sufficient that it be barrelled and each weakly closed bounded set $X \subset E$ be weakly compact.

For the proof we note that, in terms of duality, barrels in an LCS $E$ can be characterised as the polars of $w^*$-bounded41 subsets $Y \subset E^*$. Therefore any barrel in $E$ contains a neighbourhood of zero for the topology on $E$ induced by $\beta(E^{**}, E^*)$. It is clear from this that if $E$ is reflexive it must be barrelled. The second necessary condition follows from the Banach-Alaoglu theorem. Conversely, if the stated conditions are satisfied, we have $\beta(E^*, E) = \tau(E^*, E)$ and hence by the Mackey-Arens theorem, $\beta(E^*, E)$ is consistent with the duality, i.e. $E = E^{**}$. Moreover $\beta(E, E^*)$ is stronger than the initial topology on $E$ for the reasons set out above. On the other hand, any $\beta(E, E^*)$-neighbourhood of the form

$$U_{Y, \varepsilon} = \{x: |f(x)| < \varepsilon (f \in Y)\},$$

where $Y \subset E^*$ is a bounded set, contains a barrel, namely the polar of the bounded set $2/\varepsilon Y$, and so if $E$ is barrelled there is a neighbourhood of zero for the initial topology contained in $U_{Y, \varepsilon}$. \square

Thus weak compactness of the bounded weakly closed sets is a criterion for reflexivity of a barrelled LCS (in particular, a locally convex Fréchet space).

A barrelled LCS in which all the closed bounded sets are compact is called a Montel space. All finite-dimensional spaces are of this type according to the classi-
cal Borel-Lebesgue theorem. Among Banach spaces only the finite-dimensional ones are Montel spaces, but infinite-dimensional Montel space are found among Fréchet spaces, for example $C^\infty[0,1]$. The space $A(G)$ of functions which are analytic on the domain $G \subset \mathbb{C}^n$ provided with the topology of uniform convergence on compact sets is a Montel space (Weierstrass-Montel theorem). It follows from the general criterion for reflexivity that every Montel space is reflexive. Moreover the conjugate of a Montel space is a Montel space under its strong topology.

The determination of weak compactness of subsets can be difficult, because the weak topology is not in general metrizable. The following result is useful in this situation.

**Eberlein-Shmul'yan Theorem.** *If $E$ is a complete LCS and the subset $X \subset E$ is such that each sequence $\{x_n\}_1^\infty \subset X$ has a weak limit point, then its weak closure $\overline{X}$ is weakly compact.*

In the proof of this theorem we use the following criterion for $\sigma(E^*, E)$-continuity of a linear functional on $E^*$ which goes back to Banach.

**Theorem.** Let $E$ be a complete LCS. If a linear functional on $E^*$ is $w^*$-continuous on the polars of the neighbourhoods of zero in $E$, then it is $w^*$-continuous (and consequently it is generated by an element $x \in E$).

**Proof.** Let us consider the linear space $L$ of all linear functionals on $E^*$ which satisfy the condition of the theorem. Clearly $L \supseteq E$ and $(L, E^*)$ is a dual pair. We can introduce on $L$ the topology $\rho$ of uniform convergence on the polars of the neighbourhoods of zero in $E$ since the functionals in $L$ are bounded on these sets. This locally convex topology induces the initial topology on $E$ and, since $E$ is complete, it is therefore closed in $L$.

By the Banach-Alaoglu theorem the polars of the neighbourhoods of zero in $E$ are compact in the topology $\sigma(E^*, E)$. On each such set $P$ the topology $\rho$ of uniform convergence on the polars of the neighbourhoods of zero in $E$ since the functionals in $L$ are bounded on these sets. This locally convex topology induces the initial topology on $E$ and, since $E$ is complete, it is therefore closed in $L$.

By the Banach-Alaoglu theorem the polars of the neighbourhoods of zero in $E$ are compact in the topology $\sigma(E^*, E)$. On each such set $P$ the topology $\sigma(E^*, L)$ is on the one hand stronger than $\sigma(E^*, E)$, since $L \supseteq E$, while on the other hand it is weaker, since it is, by construction, the weakest topology under which all the $g|_P$ ($g \in L$) are continuous. Consequently $\rho$ is a topology of uniform convergence on a family of absolutely convex $\sigma(E^*, L)$-compact sets. By the Mackey-Arens theorem it is consistent with the duality of the pair $(L, E^*)$, i.e. $L^* = E^*$. But then if $f \in L^*$ and $f|_E = 0$ we have $f = 0$. Thus $L = E$. □

Turning to the proof of the Eberlein-Shmul'yan theorem, we stress that the non-triviality of the assertion lies in the fact that $E$ need not be weakly complete.\(^{43}\) However the space $(E^*)^*$ is complete in the $w^*$-topology and therefore $Y$, the closure of $X$ in $(E^*)^*$, is compact. It remains to show that $Y \subset E$, i.e. to establish for any functional $y \in Y$ its $w^*$-continuity on $E^*$ which, by the preceding theorem, reduces to the $w^*$-continuity of the restriction $y|_{U^o}$ on the polar of any

\(^{42}\) The topology of $C^\infty[0,1]$ is defined by the countable family of norms

\[
\|\phi\|_\infty = \max_{0 \leq k \leq n} \max_{0 \leq s \leq 1} |\phi^{(k)}(s)| \quad (n = 0, 1, 2, \ldots).
\]

\(^{43}\) For example, the Banach space $C(X)$, where $X$ is an infinite compact set, is such a space.
neighbourhood of zero \( U \subset E \). It suffices to establish continuity at zero. In the absence of this for some \( \epsilon > 0 \) we can construct inductively \( \{x_n\}_n ^\infty \subset X \) and \( \{f_m\}_m ^\infty \subset U^0 \) such that \( |f_m(z)| > \epsilon \) for all \( m \), \( |f_m(z - x_n)| < \frac{1}{3} \epsilon \) for \( m \leq n \) and \( |f_m(x_n)| < \frac{1}{3} \epsilon \) for \( m > n \).

This leads to a contradiction on taking a weak limit point \( x \) of \( \{x_n\}_n ^\infty \) and a \( w^* \)-limit point \( f \) of \( \{f_m\}_m ^\infty \).

We note further the following Shmulian-Dieudonné criterion: in a complete LCS a necessary and sufficient condition for the weak compactness of a weakly closed set \( X \) is that the intersection of any sequence \( V_1 \supseteq V_2 \supseteq \cdots \) of closed convex sets having non-empty intersection with \( X \) be non-empty.

A very general consideration of the problem of weak compactness was carried out by Grothendieck (1952) and further results were obtained by Pták (1954). To the latter is also due the credit for discovering the deep connection between the theory of duality and topological properties of homomorphisms, which are well-known as Banach theorems on the closed graph, the inverse homomorphism and the open mapping. We will discuss these fundamental results later on but for the time being we will be concerned with corresponding aspects of duality. First of all we formulate a criterion for completeness of an LCS in terms of the conjugate space.

**Theorem.** In order that an LCS \( E \) be complete, it is necessary and sufficient that a hyperplane in \( E^* \) be \( w^* \)-closed whenever its intersections with the polars of the neighbourhoods of zero in \( E \) are \( w^* \)-closed.

For the proof we associate with an arbitrary LCS \( E \) the space \( \hat{E} \) of linear functionals \( f \) on \( E^* \) for which all the sets \( f^{-1}(0) \cap U^o \) are \( w^* \)-closed (\( U \) runs through the neighbourhoods of zero in \( E \)) or, equivalently, whose restrictions \( f|_{U^o} \) are \( w^* \)-continuous. We embed \( E \) in \( \hat{E} \) in the canonical way and introduce on \( \hat{E} \) the topology of uniform convergence on the polars \( U^o \). The preceding theorem follows from the following assertion.

**Lemma.** For any LCS \( E \) the space \( \hat{E} \) is its completion, i.e. 1) \( \hat{E} \) is complete, 2) the topology of \( E \) is induced by the topology of \( \hat{E} \), 3) \( E \) is dense in \( \hat{E} \).

**Proof.** The completeness of \( \hat{E} \) follows from the \( w^* \)-compactness of each polar \( U^o \) and the completeness of the space of continuous functions \( C(U^o) \). The coincidence of the induced and the initial topologies on \( E \) is trivial. Finally the topology on \( \hat{E} \) is consistent for the dual pair \((\hat{E}, E^*)\) by the Mackey-Arens theorem, i.e. \((\hat{E})^* = E^*\). Moreover if \( E \) is not dense in \( \hat{E} \) there exists \( f \in (\hat{E})^* \) such that \( f \neq 0 \), \( f|_E = 0 \), which is a contradiction since \( f \in E^* \).

**Remark.** The criterion for completeness of an LCS can be reformulated as follows: \( E \) is complete \( \iff \) each linear functional on \( E^* \) which is \( w^* \)-continuous on the polars of the neighbourhoods of zero is \( w^* \)-continuous.

An LCS \( E \) is said to be **fully complete** (or to be a Pták space) if each subspace of \( E^* \), whose intersections with the polars of the neighbourhoods of zero in \( E \)
are $w^*$-closed, is $w^*$-closed. It follows from the previous result that a fully complete space is complete. The converse is false, however the following holds.

**Banach-Krejn-Shmul'yan Theorem.** Each locally convex Fréchet space is fully complete.

### 3.8. Continuous Homomorphisms.

Let $E_1$ and $E_2$ be LTSs. Continuity of a homomorphism $h: E_1 \to E_2$ is equivalent to its continuity at zero, for which in turn it is sufficient (and for locally bounded $E_2$ also necessary) that $h$ be bounded in the sense that the image of some neighbourhood of zero is bounded.

**Example.** In the space $C^\infty[0, 1]$ differentiation $d/dt$ is continuous but not bounded.

For any continuous homomorphism $h$ the images of the bounded sets are bounded, since if $(x_n)^\infty_1$ is a bounded sequence and $\alpha_n \to 0$, then $\alpha_n x_n \to 0$ and hence $\alpha_n (hx_n) = h(\alpha_n x_n) \to 0$. If $E_1$ is a normed space and the homomorphism $h: E_1 \to E_2$ maps bounded sets into bounded sets then $h$ is bounded and so continuous.

**Example.** Let $E$ be an LCS and $\tilde{E}$ the same space but now provided with its weak topology, which we will assume to be different from the initial topology (as a rule this is generally the case). Then the identity homomorphism $\tilde{E} \to E$ is not continuous (and so not bounded) but it takes bounded sets into bounded sets.

In the case of normed spaces $E_1$, $E_2$ boundedness of a homomorphism $h$ is equivalent to finiteness of its norm

$$
\|h\| = \sup_{\|x\| = 1} \|hx\|.
$$

In this case the set of bounded (i.e. continuous) homomorphisms of $E_1 \to E_2$ is itself a normed space (even a Banach space if $E_2$ is a Banach space).

For any LTSs $E_1$, $E_2$ the set of continuous homomorphisms of $E_1 \to E_2$ is denoted by $\text{Homc}(E_1, E_2)$. It is a linear space and can be turned into an LTS by the following standard procedure: a covering $\Omega$ of $E_1$ by bounded sets is chosen, after which a base of neighbourhoods of zero in $\text{Homc}(E_1, E_2)$ is given by the sets $N_{\nu, X} = \{h: hX \subset V\}$, where $X \in \Omega$ and $V$ is an arbitrary neighbourhood of zero in $E_2$. This is the $\Omega$-topology or the topology of uniform convergence on the family $\Omega$. It becomes stronger as the family $\Omega$ is extended but, at the same time, it is not changed if we adjoin to the sets of $\Omega$ all their subsets, finite unions and linear combinations; it is also unaffected if all the sets in $\Omega$ are replaced by their closures and if $E_1$ and $E_2$ are LCSs, all the sets in $\Omega$ can be replaced by their closed absolutely convex hulls. If $E_2$ is an LCS and $\{p_i\}_{i \in I}$ is a family of basic seminorms the $\Omega$-topology on $\text{Homc}(E_1, E_2)$ is defined by the family of seminorms $\pi_{X,i}(h) = \sup_{x \in X} p_i(hx) (X \in \Omega)$, i.e. $\text{Homc}(E_1, E_2)$ turns out to be an LCS.

---

44 It is even an algebra (associative, with unity) in the case $E_1 = E_2$. We use the notation: $\text{Homc}(E, E) = LC(E)$. 
Among the $\Omega$-topologies on $\text{Hom}_c(E_1, E_2)$ the weakest is the topology of pointwise convergence and the strongest is the topology of uniform convergence on all the bounded subsets of $E_1$. The latter is termed briefly the uniform topology. In the case $\text{Hom}_c(E, K) = E^*$ this is just the topology $\beta(E^*, E)$. For normed spaces $E_1, E_2$ the uniform topology on $\text{Hom}_c(E_1, E_2)$ is generated by a norm\(^{45}\). The topology of pointwise convergence on $\text{Hom}_c(E_1, E_2)$ is sometimes called the strong topology\(^{46}\), since the weak topology defined in the case where $E_2^*$ is total by the neighbourhoods of zero $\{h: |f(hx)| < \varepsilon\}$ ($x \in E_1, f \in E_2^*, \varepsilon > 0$) is also used (this is clearly locally convex).

The topology of uniform convergence on compact sets is also contained in the class of $\Omega$-topologies. In this case we have the well-known criterion for precompactness\(^{47}\) of a family of continuous (not necessarily linear) mappings of $E_1 \to E_2$ ($E_2$ metrizable), namely equicontinuity of the restrictions to each compact set $Q \subset E_1$ and pointwise precompactness (Arzelà-Ascoli theorem).

The following assertion follows in turn from the Banach-Steinhaus theorem (and also bears the same name\(^{48}\)).

**Theorem.** If $E_1$ is a barrelled space and $E_2$ is any LCS then each pointwise bounded family $F \subset \text{Hom}_c(E_1, E_2)$ is uniformly continuous and consequently uniformly bounded on some neighbourhood of zero.\(^{49}\)

**Proof.** Let $\{p_i\}_{i \in J}$ be a family of basic seminorms. Let us consider the continuous seminorms $p_i(hx) (h \in F)$ on $E_1$. According to the condition $\pi_i(x) = \sup_h p_i(hx) < \infty$. Consequently the seminorms $\pi_i$ are continuous and hence the uniform continuity of the family $F$ follows immediately. \(\Box\)

**Remark.** If an LCS $E_1$ is such that the Banach-Steinhaus theorem holds for $E_2 = K$ (i.e. in $E_1^*$) then $E_1$ is barrelled. In fact any barrel $T \subset E_1$ is the polar of a $w^*$-bounded and therefore equicontinuous set $F \subset E_1^*$. But then on some neighbourhood of zero $U \subset E_1$ we will have $|f(x)| < 1$ for all $f \in F$ and hence $U \subset T$.

---

\(^{45}\)If $E$ is a normed space, the norm on $L^C(E)$ satisfies the usual properties: $\|h_1, h_2\| \leq \|h_1\| \|h_2\|$, $\|1\| = 1$. Thus the continuous linear operators on $E$ form a Banach algebra $\mathcal{L}(E)$ which is noncommutative if dim $E > 1$. We note here that in any Banach algebra the operation of multiplication is continuous.

\(^{46}\)In the case of $E^*$ it coincides with the $w^*$-topology.

\(^{47}\)A set $X$ in an LTS $E$ is said to be precompact if it is relatively compact in the completion of the space $E$ (a set is said to be relatively compact if its closure is compact). The property of precompactness is weaker than relative compactness (in a complete space they are clearly equivalent), but it is convenient to work with it in view of Hausdorff's criterion: $X$ is precompact $\iff$ for any neighbourhood of zero $U$ there exists a finite $U$-net, i.e. a covering of the set $X$ by neighbourhoods of the form $x_i + U$ with $i = 1, \ldots, m$.

\(^{48}\)It is also called the principle of uniform boundedness.

\(^{49}\)If $E_2$ is a normed space uniform boundedness on some neighbourhood of zero is equivalent to uniform continuity. If both spaces $E_1, E_2$ are normed then uniform continuity of a family $F$ is equivalent to boundedness in norm: $\sup_{h \in F} \|h\| < \infty$. 
For any LTSs $E_1, E_2$ and any $h \in \text{Hom}(E_1, E_2)$ the conjugate\(^5\) homomorphism $h^*: E_2^* \to E_1^*$ is defined by the formula $(h^*f)(x) = f(hx)$ $(x \in E_1, f \in E_2^*)$. It is important to know for what topologies on the dual space $h^*$ will be continuous.

**Lemma.** Let $E_1^*$ and $E_2^*$ be provided respectively with $\Omega_1$- and $\Omega_2$-topologies where $h\Omega_1 \subset \Omega_2$. Then $h^*$ is continuous.

Verification of this basic result presents no problem. In particular, $h^*$ is continuous if $E_1^*$, $E_2^*$ are provided with their $w^*$-topologies or if both have their uniform topologies. In the latter situation if both spaces are normed then $\|h^*\| = \|h\|$. We note in passing the duality relations between the images and the kernels of the homomorphism $h \in \text{Hom}(E_1, E_2)$ and of $h^* \in \text{Hom}(E_2^*, E_1^*)$. All the annihilators considered below are relative to the pairs $(E_1, E_1^*)$, $(E_2, E_2^*)$ (not necessarily dual pairs).

With no restrictions on the LTSs $E_1$ and $E_2$ we have the formula

$$\text{Ker} h^* = (\text{Im} h)^\perp$$

and the inclusions

$$\overline{\text{Im}} h^* \subset (\text{Ker} h)^\perp, \quad \text{Ker} h \subset (\text{Im} h^*)_\perp, \quad \overline{\text{Im}} h \subset (\text{Ker} h^*)_\perp$$

(the closure of $\text{Im} h^*$ is taken in the $w^*$-topology on $E_1^*$). If $E_2^*$ is total we have

$$\overline{\text{Im}} h^* = (\text{Ker} h)^\perp, \quad \text{Ker} h = (\text{Im} h^*)_\perp, \quad \overline{\text{Im}} h = (\text{Ker} h^*)_\perp.$$

In all three cases equality is established by considering a continuous linear functional which annihilates the smaller space.

In many applications we have to deal with discontinuous homomorphisms, however the weaker property of closedness of the graph

$$\Gamma_h = \{(x, hx)\}_{x \in E_1} \subset E_1 \times E_2 \quad (h \in \text{Hom}(E_1, E_2))$$

usually holds; this is called closedness of the homomorphism itself.

**Example.** Let $E_1 = C^\infty(0, 1], E_2 = C[0, 1]$, where the topology on $C[0, 1]$ is the usual one and the topology on $C^\infty(0, 1]$ is that induced by $C[0, 1]$. Differentiation $d/dt: E_1 \to E_2$ is discontinuous but closed.

**Closed Graph Theorem (CGT).** Let $E_1, E_2$ be LCSs with $E_1$ barrelled and $E_2$ fully complete. Then if the homomorphism $h: E_1 \to E_2$ is closed it is continuous.

---

\(^5\)More precisely, the topological conjugate. Clearly $h^* = h^*|_{E_2^*}$; and moreover $\text{Im} h^* \subset E_1^*$. We also note the following formulae: 1) $(h_1 + h_2)^* = h_1^* + h_2^*$; 2) $(ah)^* = ah^*$; 3) $(hg)^* = g^*h^*$; 4) $1^* = 1$. If $h \in \text{Hom}(E_1, E_2)$ has a continuous inverse then $h^*$ is invertible and $(h^*)^{-1} = (h^{-1})^*$ but $(h^*)^{-1}$ may not be continuous.

\(^6\)If $E_1$ is a reflexive Banach space the $w^*$-topology on $E_1^*$ coincides with the weak topology of the normed space $E_1^*$ and the weak closure of any subspace of $E_1^*$ coincides with its strong closure.
In the proof of the CGT we make use of the following general criterion for closedness of a homomorphism.

**Lemma 1.** Let $E_1$, $E_2$ be LCSs and let $h \in \text{Hom}(E_1, E_2)$. A necessary and sufficient condition for closedness of $h$ is that the space $A_h$ of those $f \in E_2^*$ for which $h^*f \in E_1^*$ be dense in $E_1^*$ under the $w^*$-topology.

**Proof.** For any LTSs $E_1$, $E_2$ the conjugate space of $E_1 \times E_2$ is naturally identified with $E_1^* \times E_2^*$ in accordance with the formula $\varphi(x, y) = g(x) + f(y)$ ($g \in E_1^*$, $f \in E_2^*$). Moreover for any $h \in \text{Hom}(E_1, E_2)$ the annihilator of its graph $\Gamma_h$ is the subspace $\Gamma_h^\perp$ of functionals of the form $f(y - hx)$ where $f \in A_h$. Now let $E_1$, $E_2$ be LCSs and suppose that $h$ is closed. If $y \in E_2$, $y \neq 0$ then, since $(0, y) \notin \Gamma_h$, there is a functional $f \in A_h$ such that $f(y) \neq 0$. Conversely, if for each $y \in E_2$, $y \neq 0$ there exists a functional $f_y \in A_h$ such that $f_y(y) \neq 0$, then $\Gamma_h = \bigcap_y \text{Ker } f_y$, from which the closedness of the graph $\Gamma_h$ follows immediately. \qed

We also require the following

**Lemma 2.** If $E_1$, $E_2$ are LCSs and $E_1$ is barrelled, then for any $h \in \text{Hom}(E_1, E_2)$ the intersections of the set $A_h \subset E_2^*$ with the polars of the neighbourhoods of zero $V \subset E_2$ are $w^*$-closed.

**Proof.** If the neighbourhood $V \subset E_2$ is absolutely convex, the closure of its inverse image is a barrel in $E_1$ and therefore it contains a neighbourhood $U$ of zero in $E_1$; the $w^*$-closedness of the set $A_h \cap V^*$ follows from its representation by means of the system of inequalities $|f(hx)| \leq 1$ ($x \in U$), $|f(y)| \leq 1$ ($y \in V$). \qed

**Proof of the CGT.** By Lemma 1 the set $A_h \subset E_2^*$ is $w^*$-dense and by Lemma 2 it is $w^*$-closed since the space $E_2$ is fully complete. Consequently $A_h = E_2^*$, i.e. $h$ is weakly continuous. If $V$ is an absolutely convex neighbourhood in $E_2$, its closure $\overline{V}$ is weakly closed and therefore the inverse image $h^{-1}\overline{V} \subset E_1$ is a barrel. But then there exists a neighbourhood of zero $U$ in $E_1$ such that $U \subset h^{-1}V$, i.e. $hU \subset \overline{V}$. \qed

The following result is one of the most important consequences of the CGT.

**Inverse Homomorphism Theorem (IHT).** Let $E_1$, $E_2$ be LCSs with $E_1$ barrelled and $E_2$ fully complete. If the homomorphism $g: E_2 \to E_1$ is bijective and continuous then it is a topological isomorphism, i.e. $g^{-1}$ is continuous.

In fact the inverse of a closed (and so of a continuous) homomorphism is closed.

In a certain sense the next fundamental theorem is the dual assertion for the CGT.

**Open Mapping Theorem (OMT).** Let $E_1$, $E_2$ be LCSs with $E_1$ barrelled and $E_2$ fully complete and suppose that the homomorphism $g: E_2 \to E_1$ is surjective and continuous. Then $g$ is an open mapping.

---

52 For any LCS $E$ a subspace $N \subset E^*$ is $w^*$-dense if and only if it is total.
Proof. The space \( \text{Ker } g \) is closed\(^{53} \), therefore we can pass to the factor space \( G = E_2/\text{Ker } g \) and consider the bijective homomorphism \( \gamma : G \to E_1 \) associated with \( g \) by the relationship \( g = \gamma j \) where \( j : E_2 \to G \) is the canonical homomorphism. Full completeness is inherited by factor spaces and \( \gamma \) is continuous (because \( g \) is continuous and \( j \) is open). According to the IHT, \( \gamma^{-1} \) is continuous, i.e. \( \gamma \) is open, but then \( g \) is also open. \( \square \)

As Banach had already shown, all three theorems are true for Fréchet spaces without the assumption of local convexity. Their applications are numerous and varied (the IHT is used particularly often). The first application which we note relates to complemented subspaces.

A closed subspace \( L \) in an LTS \( E \) is said to be complemented if it has a closed complement\(^{54} \) \( M \). For example, any closed subspace of finite codimension, in particular any closed hyperplane, is of this type. In an LCS every finite-dimensional subspace is complemented. If \( E \) is a Hilbert space, any closed subspace \( L \subset E \) is complemented, the orthogonal complement

\[
L^\perp = \{ y : (x, y) = 0 \ (x \in L) \}
\]

being a closed complement for \( L \). The decomposition \( L + L^\perp = E \) (in this case we also write\(^{55} \) \( L \oplus L^\perp = E \)) is obtained by associating with each \( x \in E \) its orthogonal projection \( x' \) on \( L \) (moreover \( x - x' \) is the orthogonal projection of the vector \( x \) on \( L^\perp \)).

Lindenstrauss and Tzafriri (1971) showed that if every closed subspace of a Banach space \( E \) is complemented then \( E \) is isomorphic to a Hilbert space.

If \( L \) is a complemented subspace in an LTS \( E \) and \( M \) is its topological complement then the projection \( P \) onto \( L \) along \( M \) is closed. Thus if \( E \) is a barrelled fully complete LCS or a Fréchet space, \( P \) is continuous. If \( P \) is a continuous projection on a Banach space \( E \), \( P \neq 0 \), then \( ||P|| \geq 1 \). If \( ||P|| = 1 \) or \( P = 0 \) it is said to be an orthoprojection since in a Hilbert space this corresponds exactly to the orthogonal decomposition \( F = \text{Im } P \oplus \text{Ker } P \).

A bounded operator \( A \) on a Banach space is said to be a contraction\(^{56} \) if \( ||A|| \leq 1 \) (a uniform contraction if \( ||A|| < 1 \)). It is precisely in this case that distance is not increased: \( d(Ax, Ay) \leq d(x, y) \) (respectively \( d(Ax, Ay) \leq qd(x, y) \) where \( q < 1 \)), i.e. \( A \) is a contraction (uniform contraction) in the metric sense.

We now apply the IHT to a problem connected with metrization of an LTS. Two invariant metrics \( d_1, d_2 \) on a linear space \( E \) are said to be equivalent if they

---

\(^{53}\) For any closed homomorphism \( g \).

\(^{54}\) It is called a topological complement. In this connection we say that \( E \) decomposes into the topological direct sum of the subspace \( L \) and \( M \). This concept extends naturally to any finite and then by means of closure also to any infinite collection of subspaces. A system of subspaces is said to be complete if the closure of its sum is the whole space \( E \).

\(^{55}\) We also use the term orthogonal sum. This term extends immediately to the case of an arbitrary family of pairwise orthogonal subspaces (if the family is infinite it means the closure of the algebraic orthogonal sum).

\(^{56}\) If a contraction \( A \) is invertible and \( A^{-1} \) is also a contraction then \( A \) is an isometry. Isometries of a Hilbert space onto itself are called unitary operators.
define the same topology. We say further that \( d_1 \) is \textit{weaker} than \( d_2 \) (or \textit{subordinate to} \( d_2 \)) if the corresponding topologies are in this relationship. If the inequality \( d_1(x, y) \leq C d_2(x, y) \) holds for some constant \( C > 0 \) then \( d_1 \) is weaker than \( d_2 \).

**Theorem.** If the space \( E \) is complete under both metrics \( d_1 \) and \( d_2 \) and also \( d_1 \) is weaker than \( d_2 \) then the given metrics are equivalent.

**Proof.** The identity mapping acts continuously from the \( d_2 \)-topology into the \( d_1 \)-topology. According to the IHT we have a homeomorphism. □

Two norms \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) on a linear space \( E \) are said to be \textit{equivalent} if the corresponding metrics are equivalent. The relation of \textit{subordination} is defined correspondingly and moreover a necessary and sufficient condition for this is that the inequality \( \| x \|_1 \leq C \| x \|_2 \) be satisfied for some positive constant \( C \). Thus for equivalence of the norms it is necessary and sufficient that \( a \| x \|_2 \leq \| x \|_1 \leq b \| x \|_2 \) (\( x \in E \)) where \( a, b \) are positive constants. Any two norms on a finite-dimensional space are equivalent.

From the preceding theorem we obtain

**Corollary.** If a space \( E \) is a Banach space with respect to each of two norms, one of which is subordinate to the other, then these norms are equivalent.

Equivalent renorming is a powerful method for investigating topological properties of Banach spaces. We mention one example.

The following easily established result holds in a Hilbert space: if \( x_n \to x \) weakly and \( \| x_n \| \to \| x \| \) then \( x_n \to x \). It is very convenient to have this \textit{H-property} in a Banach space. It does hold in a uniformly convex space (V.L. Shmul'yan, 1939) but in the general case it is lacking. M.I. Kadets (1957) showed that in any separable Banach space it is possible to introduce an equivalent norm possessing the \textit{H-property}. He used this result in the solution of a problem of Banach on the homeomorphism of all infinite-dimensional separable Banach spaces.

We note that in any separable Banach space it is possible to introduce an equivalent strictly convex norm (Clarkson, 1936) and also an equivalent smooth norm (Day, 1955). The separability requirement is essential.

In 1941 Day showed that a reflexive Banach space need not have a uniformly convex equivalent norm. The possibility of introducing a uniformly convex norm is equivalent to the super-reflexivity of the space. This deep result was obtained by Enflo (1972). The concept of super-reflexivity, which is defined in the following way, is also deep. Let \( E, E_1 \) be Banach spaces. We say that \( E_1 \) is \textit{finitely representable} in \( E \) if for each finite-dimensional subspace \( L \subset E_1 \) and each \( \alpha > 1 \) there exists an embedding \( i: L \to E \) which is \( \alpha \)-close to an isometry:

\[
\alpha^{-1} \| x \| \leq \| ix \| \leq \alpha \| x \| \quad (x \in L).
\]

The space \( E \) is said to be \textit{super-reflexive} if each space which is finitely representable in it is reflexive. Super-reflexivity of \( E^* \) is equivalent to super-reflexivity of \( E \) (James, 1972). As a corollary to Enflo's theorem the complete duality between uniformly convex and uniformly smooth Banach spaces shows that the possibility
of introducing a uniformly smooth norm is also equivalent to super-reflexivity. Further in a super-reflexive space $E$ there is an equivalent norm which is simultaneously uniformly convex and uniformly smooth.

Returning after this digression to other applications of the IHT, let us consider a separable Banach space $E$ with a Schauder basis $B = (e_i)_0$. Let $\hat{E}$ be the corresponding coordinate space. By introducing the norm

$$||\xi|| = \sup_{n} \left| \sum_{i=0}^{n} \xi_i e_i \right| \quad (\xi = (\xi_i)_0)$$

on $E$ we turn it into a Banach space. The expansion in terms of $B$ can be regarded as a homomorphism $g : \hat{E} \rightarrow E$ which is clearly continuous ($||g|| = 1$) and bijective. According to the IHT, $h = g^{-1}$ is also continuous and then

$$||\xi|| \leq 2||\hat{\xi}|| \leq 2||h|| ||x||.$$ 

Thus in a Banach space the coordinates of a vector with respect to a Schauder basis are continuous linear functionals (theorem of Banach).

A Schauder basis $(e_i)_0$ in a separable Hilbert space $E$ is called a Riesz basis if the set $\hat{E}$ coincides with $l^2$. In this case

$$\sup_n \sqrt{\sum_{i=0}^{n} |\xi_i|^2} < \infty \quad (\xi \in \hat{E})$$

and, by the Banach-Steinhaus theorem, $||\hat{\xi}||_{l^2} \leq C||\hat{\xi}||_{\ell^2}$. Since $\hat{E}$ is complete under both norms they are equivalent, i.e. $1 : \hat{E} \rightarrow l^2$ is an isomorphism of the Banach spaces. Choosing an orthonormal basis $(u_i)_0$ in $E$ we consider the isometry of $l^2 \rightarrow E$ which is defined so that $\delta_i \mapsto u_i$ ($i = 0, 1, 2, \ldots$). Then the composite isomorphism $E \rightarrow \hat{E} \rightarrow l^2 \rightarrow E$ carries the Riesz basis $(e_i)_0$ onto the orthonormal basis $(u_i)_0$. Conversely, the image of an orthonormal basis under the action of any isomorphism of $E \rightarrow E$ is a Riesz basis, since under topological isomorphisms a Schauder basis carries over to a Schauder basis and the coordinate spaces are preserved. Thus we have

**Theorem of Bari.** The Riesz bases in a separable Hilbert space $E$ are the images of the orthonormal bases under all possible isomorphisms of $E \rightarrow E$.

The Riesz bases therefore form an equivalence class with respect to the action of the group of automorphisms of the Hilbert space.

**Corollary.** Any Riesz basis $(e_i)_0$ is unconditional and almost normalised in the sense that $\inf_i ||e_i|| > 0$, $\sup_i ||e_i|| < \infty$.

Toeplitz (1926) showed that any unconditional almost normalised Schauder basis in a Hilbert space is a Riesz basis.

---

57 Thus $g$ is an isomorphism of the Banach spaces $E, E$. 
Thus all unconditional almost normalised Schauder bases in a Hilbert space (which can be regarded as $l^2$) are equivalent. The situation is similar in $l^1$, in $c_0^{58}$ and (up to isomorphism) only in the stated spaces (Lindenstrauss-Zippin, 1969).

Now we make a simple but useful remark relating to the theory of bases. Let $(e_i)_{0}^{\infty}$ be a complete minimal system of vectors in a Banach space $E$. Then to each $x \in E$ there corresponds the formal expansion $x \sim \sum_{i=0}^{\infty} \xi_i(x)e_i$, where $(\xi_i)_{0}^{\infty} < E^*$ is the system which is biorthogonal to $(e_i)_{0}^{\infty}$. The partial sum $S_n = \sum_{i=0}^{n} \xi_i(\cdot)e_i$ $(n = 0, 1, 2, \ldots)$ is a bounded projection onto $\text{Lin}\{e_i\}_{0}^{\infty}$. If $(e_i)_{0}^{\infty}$ is a Schauder basis, the sequence $(S_n)_{0}^{\infty}$ converges strongly (to the identity operator) and consequently it is uniformly bounded (by the Banach-Steinhaus theorem): 

$$\sup_n \|S_n\| < \infty.$$  \hfill (18)

Conversely, it follows from (18) that $(e_i)_{0}^{\infty}$ is a Schauder basis since if $x \in \text{Lin}\{e_i\}_{0}^{\infty}$ there exists a number $v = v(x)$ such that $S_n x = x$ for all $n \geq v(x)$, i.e. $(S_n)_{0}^{\infty}$ converges to $x$ on a dense subspace and therefore also on the whole of $E$ (because of (18)).

We now mention two examples which illustrate the effectiveness of the IHT in concrete analysis. The first of these is due to Banach (1931).

**Example 1.** Let us consider the system of linear equations

$$\sum_{k=1}^{\infty} \alpha_{ik} \xi_k = \eta_i \quad (i = 1, 2, 3, \ldots).$$  \hfill (19)

Any sequence $x = (\xi_k)_{1}^{\infty}$ such that the series (19) converges with sum equal to $\eta_i$ for each $i = 1, 2, 3, \ldots$ is called a solution of the system. It turns out that, if the system (19) has a unique solution for all $(\eta_i)_{1}^{\infty}$, then the rows of the matrix $(\alpha_{ik})$ have finite support, i.e. (19) is a Toeplitz system.

For the proof we consider the space $R$ of all $x = (\xi_k)_{1}^{\infty}$ for which all the series (19) converge and introduce on $R$ a locally convex topology by means of the seminorms

$$p_i(x) = \sup_n \left| \sum_{k=1}^{n} \alpha_{ik} \xi_k \right| \quad (i = 1, 2, 3, \ldots).$$

This is a Fréchet space. The system (19) defines a continuous homomorphism $g: R \rightarrow s$, where $s$ is the space of all sequences with the topology of pointwise convergence ($s$ is also a locally convex Fréchet space). According to the condition $g$ is bijective. Consequently, $h = g^{-1}$ is continuous. Since the topology on $R$ is stronger than the pointwise topology, the linear functionals $\xi_k(\eta)$ on $s$ are continuous. But then

$$\xi_k(\eta) = \sum_{j=1}^{n_k} \beta_{kj} \eta_j \quad (\eta = (\eta_i)_{1}^{\infty}, n_k < \infty).$$

---

58 $c_0$ is the space of sequences converging to zero with the sup norm. We note that $c_0^{\infty} \approx l^1$. 
The functionals \( \xi_k(hy) \) \((y \in s)\) are linearly independent since this is true of \( \xi_k(x) \) \((x \in R, k = 1, 2, 3, \ldots)\). Consequently, the rows of the matrix \((\beta_{ik})\) are linearly independent. By the theorem of Toeplitz the system of equations

\[
\sum_{j=1}^{k} \beta_{kj} \eta_j = \xi_k \quad (k = 1, 2, 3, \ldots)
\]

has solution \( y = (\eta_j)^\circ \) for any \((\xi_k)^\circ\). But then \( \xi_k = \xi_k(hy) \) \((k = 1, 2, 3, \ldots)\) and the series

\[
\sum_{i=1}^{\infty} \delta_{ik} \eta_i \quad (i = 1, 2, 3, \ldots)
\]

converge. Because of the arbitrariness of \((\xi_k)^\circ\), the rows of the matrix \((a_{ik})\) must have finite support.

**Example 2.** Let \((M_n)^\circ\) be a sequence of positive numbers defining a quasianalytic class of functions on \([0, 1]\). Let us consider the Banach space \(B(M_n)\) of infinitely differentiable functions \(\varphi(t) (0 \leq t \leq 1)\) for which

\[
\|\varphi\| = \sup_n \frac{1}{M_n} \max_{0 \leq t \leq 1} |\varphi^{(n)}(t)| < \infty.
\]

For each such function the sequence of derivatives at zero satisfies the condition

\[
\sup_n \frac{|\varphi^{(n)}(0)|}{M_n} < \infty
\]

and \(\varphi\) is uniquely determined by \((\varphi^{(0)}(0))_0^\infty\) because of the quasianalyticity of the class. It is natural to ask if there exists for an arbitrary numerical sequence \(a = (a_n)^n_0\) satisfying the condition

\[
\|a\| = \sup_n \frac{|a_n|}{M_n} < \infty
\]

(20)

a function \(\varphi \in B(M_n)\) such that \(\varphi^{(n)}(0) = a_n \) \((n = 0, 1, 2, \ldots)\). The answer turns out to be negative except for the analytic case \(M_n = O(n!)\). For the proof we consider the Banach space \(b(M_n)\) of sequences satisfying condition (20) and the homomorphism \(g: B(M_n) \to b(M_n)\) defined by the formula \((g\varphi)_n = \varphi^{(n)}(0)\). Clearly \(g\) is continuous (\(\|g\| = 1\)), injective and, in the case where there is a positive answer to the question posed, bijective. But then \(h = g^{-1}\) is continuous. Let \(\alpha_n = o(M_n)\) \((n \to \infty)\). Then

\[
a = \sum_{n=0}^{\infty} \alpha_n \delta_n,
\]

where the series converges in \(b(M_n)\). Consequently

\[
(ha)(t) = \sum_{n=0}^{\infty} \alpha_n \frac{t^n}{n!},
\]

where the series converges in \(B(M_n)\) and so in particular at the point \(t = 1\).
Moreover

\[ \left| \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} \right| \leq C \sup_n \frac{\alpha_n}{M_n} \quad (C = M_0 \|h\|). \]

Setting \( \alpha_n = \delta_{nk} M_k \) we obtain \( M_k \leq C k! \).

The last example of the application of the IHT which we present here is concerned with the regularisation of Noetherian operators. Let \( E \) be any LTS in which the IHT holds and let \( T \) be a Noetherian operator on \( E \) whose image \( \text{Im} \ T \) is closed\(^6\) (its kernel \( \text{Ker} \ T \) is also closed since it is finite-dimensional). The regularisers for \( T \) are defined by the formulae (see Section 1.3) \( R_i = (1 - A) \tilde{T}_i P \), \( R'_i = \tilde{T}_i^2 P \), where \( A \) is a projection onto \( \text{Ker} \ T \), \( P \) is a projection onto \( \text{Im} \ T \) and \( \tilde{T}_i = T|_{\text{Ker} A} \). The projections \( A, P \) can be chosen to be continuous and \( \tilde{T}_i^{-1} \) is continuous in accordance with the IHT. Therefore both regularisers turn out to be continuous.

A sufficient condition for the preservation of the Noetherian property, normal resolvability and the value of the index under the change from \( T \) to \( T + V \), where \( V \in \mathcal{L}(E) \), is that the operators \( 1 + R_i V \) and \( 1 + VR_i \) be invertible. An effective formulation is obtained from the following general result: in any Banach algebra the elements of the form \( e - x \) with \( \|x\| < 1 \) are invertible.\(^6\) Thus in the Banach space \( E \) it is sufficient to impose on the perturbation \( V \) the condition of smallness of norm (Atkinson, I. Ts. Gokhberg, 1951), namely

\[ \|V\| < \max(\|R_i\|^{\frac{1}{2}} \|R_r^{-1}\|). \]

Invertibility of the operator \( 1 - U \) where \( \|U\| < 1 \) is a simple but very useful result. In terms of the equation \( x = Ux + y \) it implies unique solvability for all \( y \) (this also follows from the contraction mapping principle). Among its many applications there is, for example, the *theorem on stability of bases*: if \( (e_i)^{0 \infty}_i \) is a Schauder basis in a Banach space \( E \) and \( (\xi_i)^{0 \infty}_i \) is a sequence of coordinate functionals then each system of vectors \( (u_i)^{0 \infty}_i \) satisfying the condition

\[ \sum_{i=0}^{\infty} \|\xi_i\| \cdot \|e_i - u_i\| < 1 \]

is a Schauder basis in \( E \) (M.G. Krejn, D.P. Milman, M.A. Rutman, 1940). In fact the identities

\[ x = \sum_{i=0}^{\infty} \xi_i e_i, \quad Ux = \sum_{i=0}^{\infty} \xi_i (e_i - u_i) \]

define a linear operator \( U \) with norm less than 1, therefore \( S = 1 - U \) is an automorphism; but \( S e_i = u_i \) (\( i = 0, 1, 2, \ldots \)).

\(^6\) An operator \( T \) with closed image is said to be *normally resolvable*. This term is connected with the equation \( T x = y \). If \( E^* \) is total, \( \text{Im} \ T = (\text{Ker} T^*)_1 \). In this case normal resolvability is equivalent to the identity \( \text{Im} \ T = (\text{Ker} T^*)_1 \).

\(^6\) In fact, the *Neumann series* \( \sum_{k=0}^{\infty} x^k \) converges absolutely (since \( \|x^k\| \leq \|x\|^k \)) and its sum is \( (e - x)^{-1} \). We note that it follows from this result that the *set of invertible elements of a Banach algebra* is open.
Here is a further example.

**Wiener-Paley theorem.** If \((e_i)_{i=0}^\infty\) is an orthonormal basis in a Hilbert space, any system \((u_i)_{i=0}^\infty\) which satisfies the inequality
\[
\| \sum_{i=0}^n \alpha_i (e_i - u_i) \|^2 \leq \theta \sum_{i=0}^n |\alpha_i|^2 \quad (0 \leq \theta < 1).
\]
for all possible systems of numbers \((\alpha_i)_{i=0}^\infty\) is a Riesz basis.

### 3.9. Linearisation of Mappings.

Let \(E_1, E_2\) be Banach spaces and let \(G \subset E_1\) be an open set. A mapping \(F: G \rightarrow E_2\) is said to be (Fréchet) differentiable at the point \(x \in G\) if there exists a bounded homomorphism \(g: E_1 \rightarrow E_2\) such that
\[
\| F(x + v) - Fx - gv \| = o(\| v \|) \quad (v \in E_1, \| v \| \rightarrow 0).
\]
The homomorphism \(g\) is uniquely defined. It is called the (Fréchet) derivative of the mapping \(F\) at the point \(x\) and correspondingly it is denoted by \(F'(x)\) (further it can be written in the traditional way: \(dF(x) = F'(x) \, dx\)). From standard results of elementary analysis we note the formula of finite increments in the form
\[
Fx_2 - Fx_1 = \int_0^1 F'(x_1 + \tau(x_2 - x_1))(x_2 - x_1) \, d\tau.
\]
The following are conditions which are sufficient for its validity: 1) \([x_1, x_2] \subset G\); 2) \(F\) is differentiable at all points of the interval \([x_1, x_2]\); 3) the derivative \(F'(x_1 + \tau(x_2 - x_1))\) is a weakly continuous function of \(\tau\) \((0 \leq \tau \leq 1)\). They are clearly satisfied in a convex domain \(G\) for a differentiable mapping whose derivative is weakly continuous. If moreover \(q \equiv \sup_{x \in G} \| F'(x) \| < \infty\) then the mapping satisfies the Lipschitz condition
\[
\| Fx_2 - Fx_1 \| \leq q \| x_2 - x_1 \| \quad (x_1, x_2 \in G).
\]
Consequently it is uniformly continuous and extends by continuity to a mapping \(\bar{F}: \bar{G} \rightarrow E_2\). Let \(E_2 = E_1\) and suppose that \(FG \subset G, q < 1\). Then \(\bar{F}\) is a uniform contraction and consequently there exists in \(\bar{G}\) a unique fixed point \(\bar{x}\); for any \(x_0 \in G\) the iterates \(F^n x_0\) converge to \(\bar{x}\) with bound \(O(q^n)\).

For any differentiable mapping \(F: G \rightarrow E_2\) the mapping \(x \mapsto F'(x)\) is defined from the set \(G\) into the Banach space \(\text{Homc}(E_1, E_2)\). If this mapping is differentiable at some point \(x \in G\) then by definition its derivative \(F''(x)\) (the second derivative of \(F\)) belongs to the space \(\text{Homc}(E_1, \text{Homc}(E_1, E_2))\). Subsequent derivatives are defined in a similar way.

Linearisation of the mapping \(\Phi: G \rightarrow E_2\) allows us to construct the Newton process
\[
x_{n+1} = x_n - [\Phi'(x_n)]^{-1} \Phi x_n \quad (n = 0, 1, 2, \ldots)
\]
for solving the equation \(\Phi x = 0\). The convergence of this process and of certain of its modifications was investigated in detail by L.V. Kantorovich (1948). In all cases, if 1) the root \(\bar{x}\) is simple in the sense that the operator \(\Phi'(\bar{x})\) has a bounded
inverse, 2) the second derivative $\Phi''(\bar{x})$ exists and is continuous in the uniform topology on some ball $D(\bar{x}, \delta) = \{x: \|x - \bar{x}\| < \delta\}$, then for all $x_0$ in a sufficiently small ball $D$, the Newton process will be defined and will converge to the root $\bar{x}$.

In fact, the formula

$$F \bar{x} = x - [\Phi'(x)]^{-1} \Phi x$$

defines a differentiable mapping $F: D \to E_1$:

$$F'(x) dx = [F'(x)]^{-1} (F''(x) dx)[F'(x)]^{-1} F x.$$

$F'(x)$ is continuous in the uniform topology and $F'(\bar{x}) = 0$. Therefore

$$\sup_{x \in D_\varepsilon} \|F'(x)\| < 1$$

for sufficiently small $\varepsilon$.

A further standard area for applications of linearisation is the local theory of differential equations. The language of functional analysis provides a single means for considering evolutionary systems, both finite-dimensional (systems of ordinary differential equations) and infinite-dimensional (partial differential equations, integro-differential equations). An abstract evolutionary system in a Banach space $E$ has the form

$$\dot{x} = \Phi(x, t),$$

where $x = x(t)$ is a vector-function with values in $E$ defined on some interval $I$ of the real axis (we will assume further that it is on the semiaxis $t \geq 0$) and having continuous derivative $\dot{x}(t)$, and $\Phi: G \times I \to E$ where $G \subseteq E$ is, as before, some open set (for simplicity we will take $G = E$). If the mapping $\Phi$ is continuous and in each ball $D \subset E$ it satisfies a Lipschitz condition in $x$ uniformly with respect to $t$ on each finite interval, then the Cauchy problem which consists of finding a solution of the equation satisfying the initial condition $x(0) = x_0$ has a unique solution on some interval $0 < t < \varepsilon$. The proof follows the classical scheme of Picard:

$$x(t) = x_0 + \int_0^t \Phi(x(\tau), \tau) \, d\tau \quad (0 \leq t \leq \varepsilon),$$

i.e. the required solution is a fixed point of the mapping defined by the right hand side on the space of continuous vector-functions on $[0, \varepsilon]$. For sufficiently small $\varepsilon$ this mapping is a uniform contraction on a neighbourhood of the constant $x_0$.

The requirement of continuity of the mapping $\Phi$ is not satisfied (in the Banach space setting) if, for example, $\Phi$ is a differential operator. Therefore the scheme which has been described is not suitable for partial differential equations. Even for the investigation of the linear equation $\dot{x} = Ax$ where the operator $A$ is unbounded we require special methods which are connected with spectral theory (see Section 4.5).
§ 4. Theory of Operators

4.1. Compact Operators. A homomorphism \( h : E_1 \to E_2 \) of any LTSs is said to be compact or completely continuous if there exists in \( E_1 \) a neighbourhood of zero whose image is relatively compact.\(^1\) Any compact homomorphism is clearly bounded and so continuous.

The property of compactness is valuable because it is the basis of a satisfactory generalisation of Fredholm theory. Such a generalisation was effected for normed spaces by F. Riesz (1910), for LCSs by Leray (1930) and for arbitrary LTSs by Williamson (1954). The classical Fredholm theory is easily covered by these generalisations since linear integral operators are compact in the corresponding function spaces under the usual assumptions (for example, operators on \( C(S) \) (\( S \) compact) with continuous kernel or operators on \( L^2(X, \mu) \) with Hilbert-Schmidt kernel are of this type). The compact homomorphisms form in \( \text{Hom}(E_1, E_2) \) a subspace \( \mathcal{C}(E_1, E_2) \) which is invariant with respect to the natural action of \( \mathcal{L}(E_2) \times \mathcal{L}(E_1) : h \mapsto abh \ (a \in \mathcal{L}(E_2), b \in \mathcal{L}(E_1)) \). Thus \( \mathcal{C}(E, E) \equiv \mathcal{C}(E) \) is a two-sided ideal in the algebra \( \mathcal{L}(E) \).

If \( E_2 \) is complete \( \mathcal{C}(E_1, E_2) \) is closed in the uniform topology. The subspace \( \overline{\mathcal{C}}(E_1, E_2) \) of continuous, finite rank homomorphisms and consequently its uniform closure \( \overline{\mathcal{C}}(E_1, E_2) \) are contained in \( \mathcal{C}(E_1, E_2) \). There is considerable interest in the question of when \( \mathcal{C}(E_1, E_2) \) and \( \overline{\mathcal{C}}(E_1, E_2) \) coincide, i.e. of the possibility of uniform approximation of an arbitrary compact homomorphism by continuous, finite rank homomorphisms. In fact Fredholm constructed his theory of integral equations in this way by approximating their kernels by degenerate ones, i.e. by kernels of the form

\[ K_n(t, s) = \sum_{i=1}^{n} u_i(t) f_i(s). \]

Let \( E_2 \) be a barrelled LCS with a Schauder basis \( (e_i)_0^\infty \) and let \( h \) be a compact homomorphism of any LTS \( E_1 \) into \( E_2 \). Let us consider the partial sums

\[ S_n x = \sum_{i=0}^{n} \xi_i e_i \quad \left( x = \sum_{i=0}^{\infty} \xi_i e_i \right). \]

The family of linear operators \( S_n \) \( (n = 0, 1, 2, \ldots) \) is equicontinuous (by the Banach-Steinhaus theorem). Therefore if \( U \) is a neighbourhood of zero in \( E_1 \) such that \( hU \) is relatively compact, \( (S_n h)_0^\infty \) converges to \( h \) uniformly on \( U \) and consequently this convergence is uniform on each bounded set \( X \subset E_1 \). But all the \( S_n h \) are clearly of finite rank and continuous. Thus in the case under consideration \( \overline{\mathcal{C}}(E_1, E_2) = \mathcal{C}(E_1, E_2) \).

We say that an LTS \( E \) has the approximation property if the identity operator (and hence also any continuous operator) is a limit point of the set \( \overline{\mathcal{C}}(E, E) \equiv \overline{\mathcal{C}}(E) \).

---

\(^1\) A continuous mapping \( X \to Y \) of any Hausdorff topological spaces is said to be compact if there is a covering of the space \( X \) by open sets whose images in \( Y \) are relatively compact.
of finite rank operators under the topology of uniform convergence on relatively compact sets. We saw above that any barrelled space with a Schauder basis has the approximation property. We can show as above that if \( E_2 \) has the approximation property then \( \mathcal{G}(E_1, E_2) = \mathcal{C}(E_1, E_2) \) for any LTS \( E_1 \). The converse of this assertion holds in the category of Banach spaces.

In the example constructed by Enflo of a separable Banach space with no Schauder basis even the approximation property is lacking.

**Remark.** Let \( E_1, E_2 \) be reflexive Banach spaces, at least one of which has the approximation property. If the Banach space \( \text{Hom}_c(E_1, E_2) \) is reflexive (and only in this case), every continuous homomorphism of \( E_1 \to E_2 \) is compact, i.e. \( \text{Hom}_c(E_1, E_2) = \mathcal{C}(E_1, E_2) \) (Holub, 1973).

Passing directly to Fredholm theory in any LTS \( E \), we denote by \( A \) an arbitrary compact linear operator on \( E \) and fix a neighbourhood of zero \( U \) for which the image \( AU \) is relatively compact. A theory will be constructed for the operator \( T = 1 - A \), one of its basic results being that \( T \) is a **Fredholm operator** (i.e. it is Noetherian and \( \text{ind} \ T = 0 \)) and it is **normally resolvable** (i.e. \( \text{Im} \ T \) is closed). We note at once that the subspace \( \text{Ker} \ T \) is finite-dimensional since \( A|_{\text{Ker} \ T} = 1 \) and consequently \( \text{Ker} \ T \) is locally compact.

**Lemma 1.** The operator \( T \) is normally resolvable.

**Proof.** Using the compactness of \( A \), it is easy to verify that if \( X \subset \bar{U} \) and \( X \) is closed then \( TX \) is closed. Therefore the cylinder \( V = T^{-1}(TU) = \bar{U} + \text{Ker} \ T \) is closed while at the same time the cylinder \( V_0 = T^{-1}(TU) = U + \text{Ker} \ T \) is open. Consequently \( Z = V \setminus V_0 \) is a closed set. Its inflation \( \tilde{Z} = \{ x: x = az \ (a \in Z, \ a \geq 1) \} \) together with the cylinder \( V \) covers the whole space \( E \). But \( TV = T\bar{U} \) is closed. It remains to establish the closedness of the set \( T\tilde{Z} = \tilde{T}\bar{U} \). This follows from the fact that the set \( T(U \cap \tilde{V}_0) \) is closed and does not contain zero. □

**Corollary.** The subspaces \( E_n = \text{Im} \ T^n \ (n = 1, 2, 3, \ldots) \) are closed.

**The Fredholm Alternative.** Either the operator \( T \) is invertible (and in this case \( T^{-1} \) is continuous) or \( \text{Ker} \ T \neq 0 \).

**Proof.** Suppose \( \text{Ker} \ T = 0 \). Using the compactness of \( T \) it is easy to verify that the homomorphism \( S \) which is inverse to \( E \xrightarrow{T} \text{Im} \ T \) is continuous. Let us suppose that \( \text{Im} \ T \neq E \) and take \( \alpha \notin \text{Im} \ T \). We consider the \( n \)-dimensional subspace \( L_n = \text{Lin}\{T^kx\}_{k=0}^{n-1} \ (n \geq 1, L_0 = 0) \). Since \( A|_{L_n} \) is injective and maps \( L_n \cap \bar{U} \) into a compact set we have that \( L_n \cap \bar{U} \) is compact.

Because of the closedness of \( \text{Im} \ T \) and the continuity of \( S \) we can construct a neighbourhood of zero \( U' \) such that \( (L_1 + U') \cap \text{Im} \ T \subset TU \). Now we take a balanced neighbourhood of zero \( U'' \) such that \( \bar{U''} \subset U \cap U' \) and choose in \( L_n \cap \bar{U}'' \ (n \geq 1) \) any point \( x_n \notin L_{n-1} + U'' \). Clearly \( x_n \neq x_m \ (n \neq m) \).

We have \( x_n = Ty_{n-1} + \alpha_n x \), where \( y_{n-1} \in L_{n-1} \). Since \( Ty_{n-1} \in L_1 + \bar{U}'' \) then \( Ty_{n-1} \in TU \) and hence \( y_{n-1} \in U \) because \( T \) is injective. Consequently the set

---

\[ \text{Under additional assumptions about } E \text{ this is guaranteed by the inverse homomorphism theorem.} \]
\( \{ Ay_n \} \) is relatively compact. But then it must be finite since
\[
Ay_n - Ay_m \equiv -x_{n+1} (\text{mod } L_n) (n > m)
\]
and hence \( Ay_n - Ay_m \notin U'' \). As a result the set \( \{ Ay_n \} \) turns out to be finite - a contradiction. ∎

Now let us consider the root subspaces \( W_n = \text{Ker } T^n \) (\( n = 1, 2, 3, \ldots \)). Since the intersections \( W_n \cap U \) are compact, if \( W_n \neq W_{n-1} \) (\( n = 2, 3, \ldots \)) we can choose for each \( n \) a point \( x_n \in (W_n \cap U) \setminus (W_{n-1} + U) \). Then \( Ax_n - Ax_m \notin U \) (\( n > m \)) in spite of the relative compactness of \( \{ Ax_n \} \). Consequently there exists a number \( v \) such that \( W_v = W_{v+1} = \cdots (W_v \neq W_{v-1}) \). The maximal root subspace \( W_v \) is finite-dimensional.

The operator \( T \) maps the invariant subspace \( E_v = \text{Im } T^v \) into itself and moreover \( E_{v+1} \) is its image. But \( \text{Ker } T \cap E_v = 0 \). By the Fredholm alternative \( E_{v+1} = E_v \) and then \( E_n = E_v \) (\( n > v \)), in particular \( E_{2v} = E_v \). Therefore for any \( x \in E \) there exists \( y \in E \) such that \( \overline{T}y = y \), i.e. \( x = \overline{T}y \in W_v \). Thus we have established

**Lemma 2.** The following direct sum decomposition holds:
\[
E = \text{Ker } T^v + \text{Im } T^v.
\]

**Corollary.** The operator \( T \) is a Fredholm operator.

**Proof.** The codimension of the subspace \( \text{Im } T \cap \text{Im } T^v \) is clearly finite and equal to the codimension of \( \text{Im } (T|_{\text{Ker } T}) \) (since \( T|_{\text{Ker } T} \) is bijective). As a result it is equal to \( \text{dim } (\text{Ker } T) = \text{def } T \). ∎

The conjugate operator \( A^* \) of a compact operator \( A \) on a Banach space is compact (theorem of Schauder), however this is false in general even in an LCS. Nevertheless, in any LTS \( \text{def } T^* = \text{def } T \) since \( \text{Ker } T^* = (\text{Im } T)^{\perp} \) and \( \text{Im } T \) is a closed subspace with codimension equal to \( \text{def } T \).

Since \( T^n = 1 - A_n \), where \( A_n \) is a compact operator, and \( (T^n)^* = (T^*)^n \), we have \( \text{def } T^* = \text{def } T^n \) (\( n = 1, 2, 3, \ldots \)).

The property of compactness of an operator on an infinite-dimensional space is a special kind of smallness. This is evident in particular in the result that the index is stable under compact perturbations (Atkinson, I.Ts. Gokhberg, 1951).

**Theorem.** Let \( S \) be a normally resolvable Noetherian operator on an LTS \( E \) for which the IHT holds and let \( V : E \rightarrow E \) be a compact operator. Then \( S + V \) is Noetherian, normally resolvable and
\[
\text{ind } (S + V) = \text{ind } S.
\]

**Proof.** The regularisors \( R_l, R_r \) for \( S \) are continuous. In the usual notation we have
\[
R_l(S + V) = 1 + (R_lV - A), \quad (S + V)R_r = 1 + (VR_r - B).
\]
Since the parts added to the identity operator are compact we have that \( R_l(S + V) \) and \( (S + V)R_r \) are normally resolvable Fredholm operators, from which all the assertions follow. ∎
One of the most important areas of application of this theorem is the theory of singular integral equations$^3$. The kernel of a singular integral operator is frequently the sum of a standard singular kernel (for example, the Cauchy kernel $(\zeta - z)^{-1}$) and a regular$^4$ kernel which defines a compact operator. In this case the calculation of the index is reduced to that for the standard kernel.

Compactness of integral operators with regular kernels is brought about by the existence of finite rank approximations. Moreover the properties of an operator are better the more rapidly it can be approximated by finite rank operators. This observation leads us to define the approximation numbers (a-numbers) of a compact homomorphism $h: E_1 \to E_2$ of normed spaces as its deviations

$$a_{n+1}(h) = \min_{g \in \mathcal{F}_n} \| h - g \| \quad (n = 0, 1, 2, \ldots)$$

from the subspaces $\mathcal{F}_n = \mathcal{F}_n(E_1, E_2) = \{ g: g \in \mathcal{F}(E_1, E_2), \text{rk} \ g \leq n \}$. Clearly, $\| h \| = a_1(h) \geq a_2(h) \geq \ldots$ and $a_n(h) \to 0$ if and only if $h \in \mathcal{F}(E_1, E_2)$. Putting $\mathcal{G}_\infty(E_1, E_2) = \mathcal{F}(E_1, E_2)$ we define further the class $\mathcal{G}_p(E_1, E_2)$ ($0 < p < \infty$) by the requirement

$$\sum_{n=1}^{\infty} \|a_n(h)\|^p < \infty.$$

The spaces $\mathcal{G}_p(E_1, E_2)$ ($0 < p \leq \infty$) are invariant with respect to the action of $\mathcal{L}(E_2) \times \mathcal{L}(E_1)$; consequently $\mathcal{G}_p(E) \equiv \mathcal{G}_p(E, E)$ is a two-sided ideal of the algebra $\mathcal{L}(E)$.

For any $h \in \mathcal{G}_p(E_1, E_2)$ the quantity

$$N_p(h) = \left( \sum_{n=1}^{\infty} \|a_n(h)\|^p \right)^{1/p}$$

defines a quasinorm. Thus $\mathcal{G}_p(E_1, E_2)$ is a complete LTS.

If $E_1$ and $E_2$ are Hilbert spaces then $N_p$ is a norm. In this case the class $\mathcal{G}_2(E_1, E_2)$, whose elements are called Hilbert-Schmidt homomorphisms, is of special interest. Suppose that $E_1$ and $E_2$ are separable and let $(e_1^{(1)})_0^\infty$, $(e_1^{(2)})_0^\infty$ be orthonormal bases in $E_1$, $E_2$ respectively. In order that $h \in \text{Homc}(E_1, E_2)$ be a Hilbert-Schmidt homomorphism it is necessary and sufficient that

$$\sum_{i=0}^{\infty} \|he_i^{(1)}\|^2 = \sum_{i,j=0}^{\infty} |(he_i^{(1)}, e_j^{(2)})|^2 < \infty,$$

since this quantity is also $N_2$ (the Hilbert-Schmidt norm). Any integral operator on $L^2(X, \mu)$ with a Hilbert-Schmidt kernel is a Hilbert-Schmidt operator, for example, the operator on $l^2$ defined by the matrix $(a_{ij})_{j,i=0}^\infty$ such that

$$\sum_{j,i=0}^{\infty} |a_{ij}|^2 < \infty.$$

---

$^3$ In this context S.G. Mikhlin (1947) discovered the stability of the index under compact perturbations in a Hilbert space.

$^4$ Or 'weakly singular'.

One of the numerous applications of the class $\mathcal{G}_2$ in a Hilbert space $E$ is concerned with the theory of bases. Let $(e_i)_0^\infty$ be an orthonormal basis in $E$ and $(u_i)_0^\infty$ an $\omega$-linearly independent system which is \textit{quadratically close to the basis} $(e_i)_0^\infty$ in the sense that

$$\sum_{i=0}^\infty \|e_i - u_i\|^2 < \infty.$$  

Then it is a Riesz basis and it is called a \textit{Bari basis} (in honour of N.K. Bari who established the result in question in 1951). The proof is based on the fact that the operator defined on $E$ by the equations $Ae_i = e_i - u_i (i = 0, 1, 2, \ldots)$ is a Hilbert-Schmidt operator and so is compact. If $T = 1 - A$ then $Te_i = u_i (i = 0, 1, 2, \ldots)$. But $\text{Ker } T = 0$ because of the $\omega$-linear independence of the system $(u_i)_0^\infty$. Therefore $T$ is an automorphism of the Hilbert space $E$.

\textbf{Remark.} M.G. Krejn (1957) characterised a Bari basis $(u_i)_0^\infty$ by means of a certain collection of intrinsic properties, namely, 1) completeness, 2) $\omega$-linear independence, 3) quadratic closeness of the Gram matrix $((u_i, u_j))_{i,j=0}^\infty$ to the identity, i.e.

$$\sum_{i,j=0}^\infty |(u_i, u_j) - \delta_{ij}|^2 < \infty.$$  

Let $E_1, E_2$ be Banach spaces. A homomorphism $h: E_1 \to E_2$ is said to be \textit{nuclear} if it can be expanded in a series of the form

$$h = \sum_{k=1}^\infty \alpha_k f_k(\cdot)u_k,$$  

where $f_k \in E_1^*, u_k \in E_2, \|f_k\| = 1, \|u_k\| = 1$ and $\sum_k |\alpha_k| < \infty$; thus series (21) converges in the uniform operator topology so that all nuclear homomorphisms are compact. The class of nuclear homomorphisms $\mathcal{H}(E_1, E_2)$ in $\text{Hom}_c(E_1, E_2)$ is an $\Omega(E_2) \times \Omega(E_1)$-invariant subspace (in $\Omega(E)$ it is a two-sided ideal $\mathcal{R}(E)$).

The \textit{nuclear norm} in $\mathcal{H}(E_1, E_2)$ is defined as $v(h) = \inf \sum_k |\alpha_k|$ over all representations of the form (21). Clearly $\|h\| \leq v(h)$. With respect to the norm just introduced $\mathcal{H}(E_1, E_2)$ is a Banach space.

Pietsch (1963) showed that all homomorphisms of the class $\mathcal{G}_1$ are nuclear and moreover for any $p > 0$ there exists a nuclear operator on $l^1$ which does not belong to $\mathcal{G}_p$. In a Hilbert space the class of nuclear operators coincides with $\mathcal{G}_1$.

If $A: E \to E$ is a nuclear operator on a Banach space $E$ we associate with each of its representations in the form (21) $(f_n \in E^*, u_n \in E)$ the number

$$\sum_{k=1}^\infty \alpha_k f_k(u_k)$$  

which we may call the \textit{trace}\footnote{More precisely, \textit{the tensor trace}. There are several other definitions of trace which in general are not equivalent.} of the operator $A$ and denote by $\text{tr } A$ whenever this
number is independent of the representation under consideration. Grothendieck
(1955) proved that the existence of the trace for all nuclear operators on $E$ is
equivalent to the approximation property; moreover in this case the trace is a
continuous linear functional on $\mathcal{N}(E)$ (with respect to the nuclear norm $v$) and
its norm is equal to one. This applies in particular to a Banach space with a
Schauder basis. On Enflo space there exists a nuclear operator which does not
have a trace, i.e. it is such that the quantity (22) depends on the representation
of the operator in the form (21). In any Banach space all operators of class $\mathfrak{S}_1$
have a trace (Pietsch, 1963).

If a Banach space $E$ has a Schauder basis $(e_i)_{i=0}^\infty$ then corresponding to each
operator $A \in \mathfrak{L}(E)$ there is a matrix $a = (a_{ji})_{j,i=0}^\infty$ which is defined by the expansion

$$Ae_i = \sum_{j=0}^\infty a_{ji}e_j \quad (i = 0, 1, 2, \ldots).$$

We can consider its trace

$$\text{tr } a = \sum_{j=0}^\infty a_{jj}$$

whenever this series converges. If $A$ is a nuclear operator $\text{tr } a$ exists and is
equal to $\text{tr } A$ and so the trace of the matrix of the operator does not depend
on the choice of basis (G.M. Litvinov, 1979). Consequently, if the basis is
unconditional,

$$\sum_{j=0}^\infty |a_{jj}| < \infty$$

(A.S. Markus, V.I. Matsaev, 1971).

Let $A$ be a nuclear operator on a Banach space $E$ which has the approximation
property and suppose that $v(A) < 1$. Then we can define the determinant

$$\det(1 + A) = \exp\{\text{tr}[\ln(1 + A)]\},$$

where

$$\ln(1 + A) = \sum_{n=1}^\infty \frac{(-1)^n}{n} A^n.$$

The prior condition $v(A) < 1$ can be removed by the following means: whatever
the nuclear operator $A$, the function $D(\lambda) = \det(1 + \lambda A)$ of the complex variable

6. Clearly $\text{tr } A$ is a linear functional on its domain of definition.
7. In this situation we can refer to the matrix trace of the operator $A$.
8. In general, if $\varphi(\lambda)$ is an analytic function on the disk $|\lambda| < r$, for any operator $A \in \mathfrak{L}(E)$ such that
   $||A|| < r$ we put

$$\varphi(A) = \sum_{n=0}^\infty \frac{\varphi^{(n)}(0)}{n!} A^n.$$
\( \lambda \) is analytic on the disk \(|\lambda| < [v(A)]^{-1}\). It can be shown that \( D \) can be continued analytically to the whole plane, i.e. it is an entire function. It remains to take as definition: \( \det(1 + A) = D(1) \).

**Remark.** If

\[
A = \sum_{k=1}^{\infty} g_k(\cdot)v_k,
\]

where \( v_k \in E, g_k \in E^* \) and \( \sum_k \|g_k\| \cdot \|v_k\| < \infty \), then

\[
\det(1 + \lambda A) = 1 + \sum_{n=1}^{\infty} c_n \lambda^n,
\]

where \( c_n = \sum_{j_1 < \cdots < j_n} \det(g_{j_1}(v_{j_1})) \cdots \det(g_{j_n}(v_{j_n})) \) \( (n = 1, 2, 3, \ldots) \) (Koch-Grothendieck formula – a generalisation of a well-known formula from linear algebra).

**Example.** \( \det(1 + g(\cdot)v) = 1 + g(v) \).

The theorem on multiplication of determinants extends to the situation just described. As a result of this it follows that \( \det(1 + A) \neq 0 \) is a criterion for invertibility of an element \( 1 + A \) \( (A \in \mathcal{B}(E)) \) in the algebra \( \mathcal{B}(E) \) with the identity adjoined.

Continuous linear functionals on the Banach spaces \( \mathfrak{S}_p(E) \) where \( E \) is a Hilbert space can be described in terms of the trace, namely

\[
f(A) = \text{tr}(AF) \quad (A \in \mathfrak{S}_p(E)),
\]

where \( F \in \mathfrak{S}_q(E) \left( \frac{1}{p} + \frac{1}{q} = 1 \right) \), and the correspondence thus established between \( \mathfrak{S}_p(E)^* \) and \( \mathfrak{S}_q(E) \) is in fact an isometric isomorphism. For \( p = 2 \), i.e. in the space of Hilbert-Schmidt operators, we can introduce a scalar product

\[
(A, B) = \text{tr}(AB^*),
\]

with respect to which \( \mathfrak{S}_2(E) \) (with the previous norm) is a Hilbert space. In this particular case the previous formula follows from the theorem of Riesz.

The concept of a nuclear homomorphism carries over to the category of LCSs if we replace the conditions on \( f_k, u_k \) by the following: the family \( \{f_k\} \) is equi-continuous, the family \( \{u_k\} \) is bounded and its closed absolutely convex null \( U \) provides a norm on \( \hat{U} = \text{Lin} U \) with respect to which \( \hat{U} \) is a Banach space. All nuclear homomorphisms are compact. The theory of the trace outlined above extends to nuclear operators on an LCS (G.M. Litvinov, 1979). Starting from the previous definitions Grothendieck (1955) introduced the important class of nuclear spaces as those spaces for which every continuous homomorphism into a Banach space is nuclear. The class of nuclear spaces is closed with respect to completions and passage to subspaces and factor spaces (and certain other...

\(^{9}\) This is one of several equivalent definitions.
natural operations). Nuclear spaces are encountered frequently in analysis. In particular the usual spaces of infinitely differentiable and analytic functions and many natural spaces of sequences are nuclear.

It is appropriate to note that all nuclear barrelled complete LCSs are Montel spaces (consequently they are reflexive). Thus if a space is nuclear and Banach, then it must be finite-dimensional.

If \( E_1, E_2 \) are LCSs a homomorphism \( h \in \text{Hom}(E_1, E_2) \) is said to be absolutely summing if each unconditionally convergent series \( \sum_n x_n \) in \( E_1 \) is mapped to an absolutely convergent series \( \sum_n h x_n \) in \( E_2 \). If \( E_1, E_2 \) are Hilbert spaces the absolutely summing homomorphisms are precisely the Hilbert-Schmidt homomorphisms (Grothendieck, 1956). The same author also established the following deep result.

**Theorem.** If \( h: E_1 \rightarrow E_2, g: E_2 \rightarrow E_3 \) are absolutely summing homomorphisms of Banach spaces then \( gh: E_1 \rightarrow E_3 \) is nuclear.

From this we obtain immediately the Dvoretski-Rogers theorem on the existence in any infinite-dimensional Banach space \( E \) of a series of vectors which converges unconditionally but not absolutely. In fact, in the contrary case, the operator \( 1: E \rightarrow E \) is absolutely summing, therefore nuclear (\( 1 = 1^2 \)) and consequently compact.

Closely connected with the Dvoretski-Rogers theorem is Grothendieck's criterion for nuclearity of a space: in order that a locally convex Fréchet space be nuclear it is necessary and sufficient that all unconditionally convergent series of vectors in it converge absolutely. Associating with any LCS \( E \) the space \( l^1(E) \) of unconditionally convergent series and the space \( l^1\{E\} \) of absolutely convergent series, Pietsch (1963) extended this criterion to the general case. He provided these spaces with natural locally convex topologies such that the embedding \( i: l^1\{E\} \rightarrow l^1(E) \) is continuous and for nuclearity of the space \( E \) it is necessary and sufficient that \( i \) be a topological isomorphism.

The problem of the existence of a Schauder basis in a nuclear space has a rather different character than in a Banach space, since all nuclear locally convex Fréchet spaces have the approximation property. However there also exists in this class a space with no Schauder basis\(^{10}\) (N.M. Zobin, B.S. Mityagin, 1974). If there is a Schauder basis in a space of this class, then it is absolute (A.S. Dynin, B.S. Mityagin, 1960).

A nuclear LCS has certain special structure. For its description we make use of (without discussion) the general topological concept of the projective limit of a directed family of topological spaces. A projective limit of LTSs (LCSs) is an LTS (LCS) which is complete if all the spaces in the given family are complete. Any complete LCS is isomorphic to a projective limit of Banach spaces. Any complete nuclear LCS \( E \) is isomorphic to a projective limit of a certain family of Hilbert spaces. If in addition \( E \) is metrizable, the family concerned can be chosen.

\(^{10}\)We note that any nuclear locally convex Fréchet space is separable.
to be a sequence \((E_n)^\infty_0\) (with the natural ordering); in this sense \(E\) is countably-Hilbert.

Let us pursue further the question of topologising the tensor product \(E_1 \otimes E_2\) of two LCSs \(E_1, E_2\), which is closely connected with the preceding material. The algebraic object \(E_1 \otimes E_2\) is obtained by factorising the space of formal linear combinations

\[
\sum_k \lambda_k (u_k \times v_k) \quad (u_k \in E_1, v_k \in E_2)
\]

by the linear hull of the set of all combinations of the following types:

\[
(u_1 \times u_2) \times v - u_1 \times v - u_2 \times v; \quad u \times (v_1 + v_2) - u \times v_1 - u \times v_2;
\]

\[
\alpha u \times v - \alpha (u \times v); \quad u \times \alpha v - \alpha (u \times v).
\]

The canonical image of the element \(u \times v \in E_1 \times E_2\) in \(E_1 \otimes E_2\) is denoted by \(u \otimes v\). If \(E\) is a further linear space then for any bilinear mapping \(\theta: E_1 \times E_2 \to E\) there exists a unique linear mapping \(\phi: E_1 \otimes E_2 \to E\) such that the following diagram is commutative (\(j\) is the canonical mapping):

\[
\begin{array}{ccc}
E_1 \times E_2 & \overset{\theta}{\longrightarrow} & E \\
\downarrow{j} & & \downarrow{\phi} \\
E_1 \otimes E_2 & & \\
\end{array}
\]

For any two seminorms \(p^{(1)}, p^{(2)}\) on \(E_1, E_2\) respectively we define their tensor product by

\[
(p^{(1)} \otimes p^{(2)})(w) = \inf \sum \lambda_k (u_k \otimes v_k)(v_k) \quad (w \in E_1 \otimes E_2),
\]

where the infimum is taken over all decompositions of the form

\[
w = \sum_k u_k \otimes v_k.
\]

This is a seminorm on \(E_1 \otimes E_2\) which has the cross-property:

\[
(p^{(1)} \otimes p^{(2)})(u \otimes v) = p^{(1)}(u)p^{(2)}(v).
\]

If \(E_1, E_2\) are LCSs their topologies are given by families of seminorms \(\{p^{(1)}\}, \{p^{(2)}\}\) respectively. Then the family \(\{p^{(1)} \otimes p^{(2)}\}\) defines a locally convex topology on \(E_1 \otimes E_2\) which is the strongest under which the canonical mapping \(j\) is continuous\(^{11}\). The completion of the space \(E_1 \otimes E_2\) with respect to this topology is denoted by \(E_1 \hat{\otimes} E_2\) and is called the topological tensor product of the spaces \(E_1, E_2\).

**Theorem of Grothendieck.** If \(E_1, E_2\) are metrizable LCSs then each element \(w \in E_1 \hat{\otimes} E_2\) can be represented in the form

---

\(^{11}\) i.e. the space of functions of finite support \(\Phi_c(E_1 \times E_2, K)\).

\(^{12}\) There are other natural topologies on \(E_1 \otimes E_2\) but we will not discuss them here.
\[ w = \sum_{k=1}^{\infty} \alpha_k (u_k \otimes v_k), \]

where \( u_k \in E_1, \lim_{k \to \infty} u_k = 0, v_k \in E_2, \lim_{k \to \infty} v_k = 0 \) and \( \sum_{k=1}^{\infty} |\alpha_k| < \infty. \)

In this sense all elements of \( E_1 \otimes E_2 \) are nuclear.

If \( E_1, E_2 \) are normed spaces then \( E_1 \otimes E_2 \) is a Banach space with norm defined by the tensor product of the norms given in \( E_1 \) and \( E_2 \).

Particular interest attaches to the tensor product of the form \( E^* \otimes E \). In this case we have the natural homomorphism

\[ \tau \left( \sum_k f_k \otimes u_k \right) = \sum_k \xi_k(\cdot) u_k \]

from \( E^* \otimes E \) to \( \mathfrak{R}(E) \). In the Banach space case the topological tensor product \( E^* \otimes E \) is also a Banach space. Since the inequality \( \nu(wz) \leq \|w\| \) (\( \nu \) is the nuclear norm) holds for all \( w \in E^* \otimes E \), then \( \tau \) can be extended by continuity to a homomorphism \( \bar{\tau}: E^* \otimes E \to \mathfrak{R}(E) \) which is clearly surjective. Consequently

\[ \mathfrak{R}(E) \cong (E^* \otimes E) / \text{Ker} \bar{\tau}. \]

This topological isomorphism is an isometry.

\[ \text{4.2. The Fixed Point Principle. A famous theorem of Brauer asserts that if } X \text{ is a compact convex set in a finite-dimensional linear space then any continuous mapping } F: X \to X \text{ has a fixed point}^{13}. \text{ It was extended to infinite-dimensional normed spaces by Schauder (1930) and then to LCSs by A.N. Tikhonov (1935).} \]

**Theorem.** Let \( E \) be an LCS, \( X \subset E \) a non-empty closed convex set and \( F: X \to X \) a continuous mapping such that the set \( Y = FX \) is relatively compact. Then there exists a fixed point: \( x = Fx \).

**Proof.** If \( x \neq Fx \) for all \( x \in X \), there exists a neighbourhood of zero \( U \) such that \( x - Fx \notin U \) (\( x \in X \)). Let \( V \) be an absolutely convex neighbourhood of zero such that \( V + V \subset U \). We can construct a finite \( V \)-chain \( \{x_k + V\}_{1}^{n} \) on the set \( Y \). The convex hull \( \tilde{Y} \subset X \) of the set \( \{x_k\}_{1}^{n} \) is compact and lies in a subspace of dimension \( \leq n \). The mapping \( F|_{\tilde{Y}} \) is uniformly continuous. We triangulate \( \tilde{Y} \) sufficiently finely so that in each open simplex we have \( Fy' - Fy'' \in V \). Let \( \{y_i\}_{1}^{N} \) be the set of all vertices of the triangulation. We denote by \( \tilde{F}y_i \) any one of the points \( x_k \) for which \( x_k - Fy_i \in V \) and extend \( \tilde{F} \) affinely in each simplex\(^{14} \). The mapping \( \tilde{F}: \tilde{Y} \to \tilde{Y} \) is continuous. By Brauer's theorem it has a fixed point \( y \). But

\[ ^{13} \text{Since all compact convex sets of dimension } p < \infty \text{ are homeomorphic, it suffices for the proof to take } X = D^p, \text{ the closed unit ball in } \mathbb{R}^p. \text{ If } x \neq Fx \text{ for all } x \in D^p, \text{ then each ray } (1 - \tau)Fx + \tau x (\tau \geq 0, x \in D^p) \text{ intersects the sphere } S^{p-1} = \partial D^p \text{ in a uniquely determined point } \Phi x. \text{ The mapping } \Phi: D^p \to S^{p-1} \text{ is a continuous retraction: } \Phi|_{S^{p-1}} = \text{I. Its existence contradicts homology theory.} \]

\[ ^{14} \text{A mapping } \psi: M \to N \text{ of convex sets is said to be affine if } \psi((1 - \tau)x + (1 - \tau)y) = \tau \psi x + (1 - \tau) \psi y (0 \leq \tau \leq 1). \]
this leads to the contradiction

$$y - Fy = \sum_{j=1}^{m} \tau_j(Fy_{ij} - Fy_{ij}) + \sum_{j=1}^{m} \tau_j(Fy_{ij} - Fy) \in V + V \subseteq U$$

for some $\tau_j > 0 \left( \sum_j \tau_j = 1 \right)$. \( \square \)

The **fixed point principle** has many important and interesting applications. We will go into the details of some of these here.

**Example 1.** Let us consider a Banach space $E$ which is ordered by a closed cone $K$. Let $T: E \to E$ be a compact (in general non-linear) mapping which is monotone in the usual sense, i.e. $x \leq y \Rightarrow Tx \leq Ty$, and positively homogeneous: $T(\alpha x) = \alpha Tx \ (\alpha > 0, x \in E)$. If under these conditions there exists a vector $v \geq 0 (\|v\| = 1)$ such that $Tv \geq cv$ for some $c > 0$, then there exists an eigenvector $e \geq 0 \ (e \neq 0)$ corresponding to some positive eigenvalue $\lambda$. This **theorem of Rutman** is a generalisation of the classical **Perron-Frobenius theorem** concerning a matrix with non-negative elements. For the proof we consider the closed convex subset $X_e$ by imposing the additional conditions: $\|x\| < 1, f(x) \geq ef(x)$. It is non-empty: if $x \in K_e$ and $\|x\| = \frac{1}{2}$ then $x \in X_e$ (in particular, $v \in X_e$). If $x \in X_e$ then $\|x\| = \frac{1}{2}$ and consequently $Tx \neq 0$ for $x \in X_e$. The formula

$$F_x = \frac{Tx + 2\|x\|v}{\|Tx + 2\|x\|v\|}$$

defines a continuous mapping of the set $X_e$ into a relatively compact subset of it (because of the compactness of $T$). By the fixed point principle there exists $x_e \in X_e$ such that $F_x x_e = x_e$, hence $\|x_e\| = 1$ and $Tx_e = \lambda x_e - \mu x$, where $\lambda_e, \mu_e > 0$. Using the compactness of $T$ once again, we choose $\varepsilon_n \downarrow 0$ such that sequence $(Tx_n)_{n=1}^{\infty}$ converges. Then $(x_n)_{n=1}^{\infty}$ will also converge to some $x \geq 0 (\|x\| = 1)$ and $(\lambda_n)_{n=1}^{\infty}$ will converge to some $\lambda \geq c$. We obtain in the limit $Tx = \lambda x$.

**Example 2.** Let us consider the problem of the existence of a non-trivial closed invariant subspace for a continuous linear operator $A$ on an LTS $E$. Using the algebraic construction (see Section 1.2) with subsequent closure does not generally give the required result since the possibility that $Lin\{A^k x\}_{k=1}^{\infty} = E$ for all $x \in E$ is not excluded (however in this case $E$ is separable). In 1984 Read constructed a counterexample in a certain non-reflexive Banach space. However, if $A$ is a compact operator on an arbitrary Banach space, it has a non-trivial closed invariant subspace (Aronszajn-Smith, 1954).

\[1^4\] The concepts of eigenvalue and eigenvector for non-linear mappings are defined exactly as for linear mappings: $Te = \lambda e$.

\[1^5\] The corresponding generalisation for integral operators was obtained by Yentsch (1912).

\[1^7\] The fact that the Perron-Frobenius theorem can be obtained by the method of fixed points (in fact from Brauer's theorem) was noted by P.S. Aleksandrov and Hopf (1935).
Does there exist such a subspace for a pair of commuting compact operators? Using the fixed point principle, V.I. Lomonosov (1973) obtained the following general result$^{18}$.

**Theorem.** Let $A$ be a non-zero compact operator on an infinite-dimensional LCS $E$. Then the algebra $(A)'$ of all continuous operators which commute with $A$ is (topologically) reducible, i.e. there is a non-trivial closed subspace which is invariant for all $T \in (A)'$.

The following lemma, which also has an important independent significance, constitutes a fundamental step of the proof.

**Lemma.** Let $\mathcal{A}$ be an irreducible subalgebra of the algebra of continuous operators on an LCS $E$ which contains a compact operator $A \neq 0$. Then there is a compact operator $B \in \mathcal{A}$ which has a fixed point $x \neq 0$.

**Proof.** For any $y \in E$, $y \neq 0$, the subspace $L_y = \{ z : z = Ty (T \in \mathcal{A}) \}$ is different from the zero subspace, invariant for all $S \in \mathcal{A}$ and, consequently, dense. We choose a point $x_0$ such that $Ax_0 \neq 0$ and a convex neighbourhood of zero $U$ such that the set $Q = Ax_0 + AU$ is compact and does not contain zero (clearly, $Q$ is convex). Then for each $y \in Q$ there is an operator $T_y \in \mathcal{A}$ such that $T_y x_0 \neq 0$. By continuity $T_y v - x_0 \in U$ for all $v \in Q$ close to $y$. There exists therefore a finite collection $\{ T_i \}_{i=1}^n \subset \mathcal{A}$ such that for each point $v \in Q$ at least one of the inclusions $T_i v - x_0 \in U$ holds, i.e. at least one of the inequalities $q(T_i v - x_0) < 1$ holds, where $q$ is the gauge of the neighbourhood $U$. We put

$$\phi(v) = \max(1 - q(v, 0), 0),$$

$$\phi_i(v) = \frac{q(T_i v - x_0)}{\sum_{j=1}^n \phi(q(T_j v - x_0))} \quad (1 \leq i \leq n).$$

The mapping

$$\Phi v = \sum_{i=1}^n \phi_i(v) T_i v \quad (v \in Q)$$

is continuous and so $\Phi Q$ is compact; moreover $\Phi Q \subset x_0 + U$ since if $T_i v \neq x_0 + U$ then $\phi_i(v) = 0$. The mapping $\Phi = \Phi A$ has a fixed point $x$ in $x_0 + U$. It is fixed for the compact operator

$$B = \sum_{i=1}^n \phi_i(Ax) T_i A \in \mathcal{A}$$

and $x \neq 0$ since $0 \notin x_0 + U$. $\square$

---

$^{18}$ For Banach spaces; however the extension to LCSs does not require significant changes.

$^{19}$(A)' is called the **centralizer** (in general algebraic terminology) or the **commutant** (in operator theory terminology) of the operator $A$. The **centraliser (commutant)** $M'$ of any set $M \subset LC(E)$ is defined as the intersection of all $(A)' \ (A \in M)$. If $\mathcal{A} \subset \mathcal{L}(E)$ is a subalgebra then $\mathcal{A} \cap \mathcal{A}'$ is the centre of the algebra $\mathcal{A}$. In any case it contains the scalar operators $A = \lambda 1 \ (\lambda \in K)$.

$^{20}$ It is said to be **hyperinvariant** for $A$. 
It is sufficient to prove the theorem of Lomonsov in the complex case. If $(A)'$ is irreducible, by the Lemma there is a compact $B \in (A)'$ for which $L \equiv \text{Ker}(1 - B) \neq 0$. Since $L$ is finite dimensional and invariant for $A$ then $A|_L$ has an eigenvalue $\lambda$. The eigenspace $\text{Ker}(A - \lambda I)$ is invariant for $T \in (A)'$ and consequently it coincides with $E$, i.e. $A = \lambda I$, hence $A = 0$ since it is compact and $E$ is infinite dimensional. 

**Remark.** The infinite-dimensionality of $E$ is not used until the last step. Therefore if we require that $A$ is not scalar\(^2\), the assertion of the theorem will also be true in the finite dimensional case but of course this case allows a purely algebraic treatment: a classical theorem of Burnside asserts that any irreducible subalgebra of the algebra $\text{End} \ E$ with finite dimensional $E$ coincides with $\text{End} \ E$. Therefore if $(A)'$ is irreducible, then $A$ commutes with all operators in $E$. But then, by Schur's lemma, $A$ is scalar.

**Corollary 1.** If $T$ is a continuous operator on an LCS\(^2\) $E$ and the algebra $(T)'$ contains a compact operator $A \neq 0$, then $T$ has a non-trivial closed invariant subspace.

**Corollary 2\(^3\).** Let $\mathfrak{A}$ be an irreducible algebra of operators on an LCS $E$ which contains a compact operator $A \neq 0$. Then its centraliser $\mathfrak{A}$ consists of the scalar operators.

**Example 3 (Markov-Kakutani theorem).** If $E$ is an LCS, $X \subseteq E$ is a compact convex set and $\{A_i\}_{i \in I}$ is any commuting family of continuous affine mappings of $X \to X$, then there exists a common fixed point $x \in X$ for all $A_i$.

In fact, for each $A_i$ its set of fixed points $X_i$ is non-empty, convex, compact and invariant for all $A_j$. By induction these properties carry over to finite intersections $X_{i_1} \cap \cdots \cap X_{i_n}$. Consequently the intersection of all the $X_i$ is non-empty.

One of the applications of the Markov-Kakutani theorem is concerned with invariant means. Let us define this important concept.

Let $S$ be any semigroup (to be specific we take it to be multiplicative). Any non-negative linear functional $m$ on the space $B(S)$ which is invariant with respect to left shifts ($m[\phi] = m[\phi]$, where $\phi(s) = \phi(ts)$) and normalised in the sense that $m[1] = 1$ is called a left-invariant mean on $S$. Clearly, the value of $m[\phi]$ is contained between inf $\phi$ and sup $\phi$ from which it follows that the functional $m$ is continuous ($\|m\| = 1$)\(^2\). If a left-invariant mean exists\(^2\) on $S$ then $S$ is said to be left-amenable. If $S$ is a group and $m$ is a left-invariant mean, by putting $\phi^*(s) = \phi(s^{-1})$ and $m^*[\phi] = m[\phi^*]$ we obtain a right-invariant mean. Thus for groups (and clearly also for Abelian semigroups) we can speak simply of amenability.
bility. Invariant means were brought into use by von Neumann (1929) who established the amenability of abelian groups and gave a simple example of a non-amenable group: the free group with two generators.

**Theorem.** Any Abelian semigroup is amenable.

**Proof.** In the space $B(S)^{**}$ the set

$$\Delta = \{ f : f(1) = 1, \| f \| \leq 1 \} = \{ f : f(1) = 1, f \geq 0 \}$$

is $w^{*}$-compact by the Banach-Alaoglu theorem, convex and invariant with respect to the mappings which are the conjugates of shifts: $(A_{t}, f) \phi = f(\phi_{t})$. Clearly all the $A_{t}$ are affine and $w^{*}$-continuous on $\Delta$ and $A_{t_{1}}A_{t_{2}} = A_{t_{1}t_{2}}$, i.e. the family $\{ A_{t} \}$ commutes since the semigroup is abelian. By the Markov-Kakutani theorem there is a common fixed point $m$ for all the $A_{t}$. This is also an invariant mean. □

An invariant mean on the additive semigroup of natural numbers $\mathbb{N}$ is called a Banach limit. This is a non-negative linear functional on the space of bounded sequences which is equal to 1 at $(1, 1, \ldots)$ and invariant with respect to the shift $(\alpha_{n}) \rightarrow (\alpha_{n+1})$. Its value at a convergent sequence is the same as the limit. A Banach limit provides a regular method for the summation of series with bounded partial sums.

### 4.3. Actions and Representations of Semigroups

Let $S$ be a topological semigroup and $X$ a Hausdorff topological space. An action of $S$ on $X$ (or in $X$) is a family of continuous mappings $T(s): X \rightarrow X$ ($s \in S$) such that $T(st) = T(s)T(t)$ and the orbit $T(s)x$ is continuous in $s$ for each fixed $x \in X$. If there is an identity $e$ in $S$ then $T(e)$ is idempotent. If $T(e) = 1$, the action is said to be monoidal. This condition is usually assumed to be satisfied if $S$ is a group and consequently in this case all the $T(s)$ are bijective.

The actions of a semigroup $S$ on a space $X$ are also called (topological) dynamical systems on $X$; in this context $X$ is called the phase space, the points $x \in X$ are the states of the system and the variable $s \in S$ is time. The following is the classical situation: $S = \mathbb{R}$ (flow), $S = \mathbb{R}^{+}$ (semiflow), $S = \mathbb{N}$ (iterations of a mapping $F: X \rightarrow X$).

One of the central problems of topological dynamics is the existence of an invariant measure, i.e. a measure $\mu$ on $X$ such that

$$\int_X \phi(T(s)x) \, d\mu = \int_X \phi(x) \, d\mu$$

for all $\phi \in L^1(X, \mu)$.

---

26 i.e. a semigroup provided with a Hausdorff topology with respect to which multiplication is continuous.

27 A semigroup with an identity is sometimes called a monoid.

28 In a topological group not only multiplication but also inversion $s \mapsto s^{-1}$ is continuous (by definition).

29 For the sake of brevity we sometimes also call them flows.
Bogolyubov-Krylov Theorem. If the phase space $X$ of a dynamical system is compact and the time semigroup $S$ is left-amenable (in particular, Abelian), there exists an invariant measure $\mu$ on $X$ with $\mu(X) = 1$.

**Proof.** Let $m$ be a left-invariant mean on $S$. We take any point $x_0 \in X$ and consider the linear functional $f(\phi) = m[\phi(T(\cdot)x_0)]$ on $C(X)$. Since it is non-negative, by the theorem of Riesz it can be represented in the form $\int_X \phi \, d\mu$. This measure $\mu$ has the required properties. \(\square\)

Any semigroup $S$ acts on itself ($X = S$) by left shifts: $T(s)x = sx$. Therefore if $S$ is a compact abelian semigroup there is a measure on it which is invariant with respect to shifts. For groups a considerably more general result holds.

**Theorem of Haar.** On any locally compact group there is a left invariant measure (left Haar measure).

In the same way there also exists a right Haar measure. Each of them is unique up to a positive multiple (von Neumann, 1936) but a left Haar measure need not be a right Haar measure. Groups for which left and right Haar measures are the same are called unimodular. In particular, all compact and all discrete groups are of this type.

Haar measure is a very valuable tool for constructing the theory of representations of locally compact groups, in particular for the construction of general harmonic analysis.

By a representation of a topological semigroup $S$ in an LTS $E$ we mean a linear action $T$ on $E$, i.e. a strongly continuous homomorphism $T: S \to \mathcal{L}(E)$. Thus all the $T(s)$ are continuous linear operators on $E$ and $T(s)x$ is a continuous function of $s$ for each $x \in E$; $T(s_1s_2) = T(s_1)T(s_2)$. If $T(s)$ depends continuously on $s$ in the uniform topology then the representation $T$ is said to be uniformly continuous.

A representation is said to be bounded if all its orbits are bounded. For a Banach space $E$ this is equivalent to boundedness in norm: $\sup_s \|T(s)\| < \infty$. If $\|T(s)\| \leq 1$ ($s \in S$) the representation is said to be contracting and if all the $T(s)$ are isometries it is called isometric. Any contracting representation of a group is isometric.

Any bounded representation $T$ of a semigroup $S$ in a Banach space is contracting in the equivalent norm

$$\|x\|_1 = \max \left( \|x\|, \sup_s \|T(s)x\| \right).$$

A representation of a group $G$ in a Hilbert space is said to be unitary if all the operators $T(g)$ ($g \in G$) are unitary.

**Theorem.** Let $S$ be a right-amenable topological semigroup and $T$ a bounded representation of it in a Hilbert space $E$. If

\footnote{In a somewhat generalised form.}
then there exists on $E$ an equivalent Hilbert norm with respect to which $T$ is unitary.

Proof. We consider on $S$ the bounded function $\phi_{x,y}(s) = (T(s)x, T(s)y)$ ($x, y \in E$) and put $\langle x, y \rangle = m[\phi_{x,y}]$. Then $\langle T(s)x, T(s)y \rangle = \langle x, y \rangle$ ($s \in S$) and $\alpha^2 \|x\|^2 \leq \langle x, x \rangle \leq \beta^2 \|x\|^2$ ($\beta = \sup_{s} \|T(s)\|$).

In many situations this theorem allows us to restrict attention to unitary representations. The following are classical examples: 1) finite-dimensional representations of finite groups, whose theory was developed at the end of the 19th and the beginning of the 20th centuries (Frobenius, Molin, Burnside, Schur); 2) representations of compact groups in Hilbert space (the theory of unitary representations of compact groups was essentially constructed in the work of Peter and H. Weyl, 1927).

If $T$ is a finite-dimensional representation of a semigroup $S$, the function $\chi(s) = \text{tr} T(s)$ ($s \in S$) is called its character. This term reflects the fundamental classical result that, for example, unitary representations with identical characters are indistinguishable from the geometric point of view; more precisely, they are equivalent in the sense defined below.

Let $T_1, T_2$ be representations of a topological semigroup $S$ in LTSs $E_1, E_2$ respectively. Any continuous homomorphism $V: E_1 \to E_2$ which satisfies the relation $T_2(s)V = VT_1(s)$ ($s \in S$) is called an intertwining operator for the pair of representations $T_1, T_2$. For example, $V = 0$ is such an operator. If among the intertwining operators there is a topological isomorphism $U$ (so that $T_2(s) = UT_1(s)U^{-1}$) we say that the representations $T_1, T_2$ are equivalent. If the space is a Hilbert space and $U$ is an isometry we use the term unitary equivalence.

In a series of situations generalisations of the concept of trace allow us to consider characters of infinite-dimensional representations and to apply them for classification purposes. On the other hand, the characters of one-dimensional representations are very important and interesting (especially in the Abelian case). They are the continuous homomorphisms of a semigroup $S$ into the multiplicative group of the field $\mathbb{C}$ and are called (one-dimensional) characters of the semigroup $S$. For example, the exponentials are the characters of the additive semigroup $\mathbb{R}$, (the character which is identically zero is usually excluded from consideration).

Now let $G$ be a locally compact group and $\mu$ a right Haar measure. The group $G$ acts by right shifts on the Hilbert space $L^2(G, \mu)$. This unitary representation $R$ of the group $G$ is said to be regular\(^{31}\) and is the starting point of harmonic analysis on a group. If $G = \mathbb{R}$ or $G = T$ (the unit circle) then $\mu$ is Lebesgue measure and harmonic analysis on $G$ is Fourier analysis. We now pass to a

---

\(^{31}\) This term is retained for representations by means of shifts in any function spaces on a group which are invariant with respect to shifts.
consideration of the basic aspects of the spectral theory of operators with which harmonic analysis has a deep connection.

4.4. The Spectrum and Resolvent of a Linear Operator. Many applications require that spectral theory covers linear operators which are not necessarily defined on the whole of the underlying LTS $E$ and are not necessarily continuous. The domain of definition $D(A)$ of an operator $A$ on $E$ can be any subspace of $E$ and in general it is neither closed nor dense, i.e. $A$ is simply a homomorphism of $D(A) \to E$. Such a generality has significant implications for the algebra of linear operators. For example $D(A + B) = D(A) \cap D(B)$ and it can happen that $D(A + B) = 0$ for dense $D(A)$, $D(B)$. We note here that $D(AB) = B^{-1}D(A)$, the $B$-preimage of the subspace $D(A)$.

In accordance with the usual formal rules, just one change in the domain of definition changes the operator. If $D(A) \subseteq D(\tilde{A})$ and $A|_{D(A)} = \tilde{A}$ then $\tilde{A}$ is called an extension of the operator $A$ and we may write $A \subset \tilde{A}$ (in this connection the operator $A$ is said to be a restriction or part of the operator $\tilde{A}$).

Let $A$, $B$ be operators such that $BA \subset AB$ and $B$ is defined on the whole space $E$. Then we say that they commute even if $BA \not= AB$, since if $BA = AB$ and $D(AB) = E$ then $D(A) = E$, a restriction which is unacceptable for many applications.

Example. Suppose that $D(A) = D(B) = E$ and $A$, $B$ commute ($AB = BA$). If $A$ is injective, the operator $A^{-1}$ ($D(A^{-1}) = \text{Im } A$) commutes with $B$ in the sense defined above.

The theory of operators is concerned in the first place with their properties which are invariant with respect to automorphisms of the underlying space. From this point of view the operator $UAU^{-1}$ ($U$ is a topological automorphism) is just as suitable as $A$; in this connection, just as for representations, we say that they are equivalent (the term similar operators is also used). In the finite-dimensional case the problem of classifying linear operators up to similarity reduces to the theory of elementary divisors. In the infinite-dimensional case the problem has on the whole a transcendental nature. However knowledge of certain, if not all, invariants already allows the possibility of penetrating deeply into the internal structure of an operator. The spectrum is the most important invariant of this sort.

As the ground field in spectral theory it is expedient both from the internal point of view and in connection with applications to settle on the field $\mathbb{C}$, which has a unique combination of properties: algebraic closure, local compactness, connectedness.

A complex number $\lambda$ is said to be a regular point of a linear operator $A$ on an LTS $E$ if $A - \lambda 1$ maps $D(A)$ bijectively onto $E$ and the resolvent $R_\lambda = R_\lambda(A) = (A - \lambda 1)^{-1}$ is a continuous operator on $E \to D(A)$ (in this connection it is convenient to regard $R_\lambda(A)$ as an operator on $E \to E$ by using the inclusion of $D(A)$ in $E$; in any event $\text{Im } R_\lambda(A) = D(A)$).

The set reg $A$ of regular points of an operator $A$ is called the resolvent set and its complement spec $A$ is the spectrum.
The spectrum of an operator contains the set of all eigenvalues and sometimes reduces to this, for example, if \( A : E \to E \) is a continuous algebraic\(^{32} \) operator.

**Example.** Let \( P \) be a projection different from 0, 1. Then the set of its eigenvalues is \( \{0, 1\} \) and if \( P \) is continuous \( \text{spec} P = \{0, 1\} \) (otherwise, \( \text{spec} P = \mathbb{C} \)).

An eigenvalue \( \lambda \) is said to be **normal** if corresponding to it there exists a finite-dimensional root subspace with an invariant topological complement \( L \) and \( \lambda \notin \text{spec}(A|_L) \).

For any linear operator \( A \) and all \( \lambda \in \mathbb{C} \) we put \( \Delta_\lambda \equiv \Delta_\lambda(A) = \text{Im}(A - \lambda I) \). The eigenvalues \( \lambda \) are characterised by the fact that the homomorphism \( (A - \lambda I) : D(A) \to \Delta_\lambda(A) \) is not invertible. For all other \( \lambda \) the inverse homomorphism is defined. As before we can call it the **resolvent** of the operator \( A \), denote it by \( R_\lambda \equiv R_\lambda(A) \) and regard it as an operator on \( E \) with domain of definition \( \Delta_\lambda(A) ; \text{Im} R_\lambda(A) = D(A) \).

The eigenvalues of an operator \( A \) are contained in the **approximate spectrum** \( \text{spec}_a A \) which is defined in the following way: \( \lambda \in \text{spec}_a A \) if there is a neighbourhood of zero \( V \) such that for any neighbourhood of zero \( U \) a vector \( x \notin V \) can be found for which \( Ax - \lambda x \in U \).

The set \( \text{spec}_r A = \text{spec} A \setminus \text{spec}_a A \) is called the **residual spectrum**. The union \( \text{reg} A = \text{reg} A \cup \text{spec}_r A \) is characterised by the fact that the resolvent \( R_\lambda : \Delta_\lambda \to E \) is a continuous operator (but \( \Delta_\lambda \neq E \) for \( \lambda \in \text{spec}_r A \) in contrast to the situation for regular points). Points \( \lambda \in \text{reg} A \) (i.e. \( \lambda \notin \text{spec}_r A \)) are said to be **quasiregular**.

If an operator \( A \) on a complete LTS \( E \) is closed\(^{33} \), the subspace \( \Delta_\lambda(A) \) is closed for all \( \lambda \in \text{reg} A \). In this case the factor space \( E/\Delta_\lambda \) is called the **defect space** and its topological dimension \( n_\lambda \) is called the **defect number** of the operator \( A \) at the point \( \lambda \).

The approximate spectrum of an operator \( A \) on a normed space \( E \) is the set of those \( \lambda \in \mathbb{C} \) for each of which there exists a **quasieigensequence** of vectors \( (x_n)_{n=1}^\infty \):

\[
\lim_{n \to \infty} (Ax_n - \lambda x_n) = 0, \quad \lim_{n \to \infty} \|x_n\| > 0.
\]

We can assume moreover that \( \|x_n\| = 1 \) \((n = 1, 2, 3, \ldots) \). The quasiregular points \( \lambda \) can be characterised by an inequality of the form

\[
\|Ax - \lambda x\| \geq c_\lambda \|x\| \quad (x \in D(A), c_\lambda > 0).
\]

The set \( \text{reg} A \) is therefore open (if \( |\mu - \lambda| < c_\lambda \) we can put \( c_\mu = c_\lambda - |\mu - \lambda| \)) and consequently the approximate spectrum is closed.

Using the technique of gaps, M.G. Krejn, M.A. Krasnosel'skij and D.P. Mil'man (1948) obtained the following general result.

**Theorem.** Let \( A \) be a closed operator on a Banach space \( E \). Then the defect number \( n_\lambda \) is constant in a sufficiently small neighbourhood of any quasiregular point.

---

\(^{32}\) The requirement of continuity is essential; the definition of the spectrum in an LTS takes topology into account.

\(^{33}\) This is guaranteed if \( \text{reg} A \neq \emptyset \).
For the proof it is sufficient to note that if \( y = Ax - \lambda x = (Ax - \mu x) + (\lambda - \mu)R_\lambda y \) and \( \|y\| = 1 \) then \( d(y, \Delta \mu) \leq |\mu - \lambda|c_\lambda^{-1} \). Therefore we have the bound for the gap

\[
\vartheta(\Delta_\lambda, \Delta_\mu) \leq |\mu - \lambda| \max \left( \frac{1}{c_\lambda |c_\lambda - |\mu - \lambda|} \right) < \frac{1}{2}
\]

if \( |\mu - \lambda| < \frac{1}{2}c_\lambda \). But then \( n_\mu = \dim \text{rk}(E/\Delta_\mu) = \dim \text{rk}(E/\Delta_\lambda) = n_\lambda \). \( \square \)

**Corollary 1.** The defect number \( n_\lambda \) is constant on each connected component of the set \( \text{reg} A \).

In particular, if \( \text{reg} A \) is connected, then either \( \text{reg} A = \text{reg} A \) or \( \text{reg} A = \text{spec} A \).

**Corollary 2.** The resolvent set \( \text{reg} A \) and the residual spectrum \( \text{spec}_r A \) are open.

In fact, these sets are separated out from \( \text{reg} A \) by the respective conditions \( n_\lambda = 0 \) and \( n_\lambda \neq 0 \).

**Corollary 3.** The spectrum of an operator \( A \) is closed.

However it can turn out to be empty.

**Example.** For the differentiation operator on \( L^2(0, 1) \) the natural (maximal) domain of definition consists of all absolutely continuous functions with derivative in \( L^2(0, 1) \). Let us restrict the domain of definition by imposing the condition \( \phi(0) = 0 \). We obtain a closed operator with no spectrum since its resolvent

\[
(R_\lambda \psi)(s) = \int_0^s e^{s(t-0)}\psi(t) \, dt
\]

is a bounded operator on the whole of \( L^2(0, 1) \) for all \( \lambda \in \mathbb{C} \).

We note further that any non-empty closed set \( S \subset \mathbb{C} \) is the spectrum of some closed linear operator. For this we consider the Hilbert space \( L^2(S, \mu) \), where \( \mu \) is a measure on \( S \) which is positive on all open subsets \( X \) of \( S \). Let us denote by \( A \) the operator obtained by multiplying by the independent variable \( \zeta : (A\phi)(\zeta) = \zeta \phi(\zeta) \). Its domain of definition consists of those \( \phi \in L^2(S, \mu) \) for which \( \zeta \phi(\zeta) \) belongs to the same \( L^2 \). The operator \( A \) is closed and if \( S \) is compact (and only in this case) then \( A \) is even bounded. It is easy to see that \( \text{spec} A = \text{spec}_r A = S \). The points with positive measure are the eigenvalues in this example.

For any operator \( A \) the resolvent \( R_\lambda \) is an analytic function of \( \lambda \) on \( \text{reg} A \) since the power series

\[
\sum_{n=0}^{\infty} R_\mu^{n+1}(\lambda - \mu)^n
\]

converges in the uniform operator topology for \( |\lambda - \mu| < \|R_\mu\|^{-1} \) and its sum is\(^{34} R_\lambda \).

\(^{34}\)This argument shows once again that \( \text{reg} A \) is open. Analyticity of the resolvent is the basis for the application of the methods of complex analysis in spectral theory.
Suppose that the operator $A$ is bounded and $D(A) = E$. The Laurent series

$$
- \sum_{n=0}^{\infty} A^n \lambda^{-n}
$$

converges in the domain $|\lambda| > \|A\|$ under the uniform operator topology and its sum is $R_\lambda$. Thus $\text{spec } A$ lies in the disk $|\lambda| \leq \|A\|$. We see that the spectrum of a bounded operator is compact. On the other hand $\text{spec } A \neq \emptyset$ since otherwise $R_\lambda$ is an entire function which tends to zero as $|\lambda| \to \infty$ and so, by Liouville's theorem, $R_\lambda \equiv 0$ which is impossible for an inverse operator.

The smallest $\rho = \rho(A) \geq 0$ for which $\text{spec } A$ lies in the disk $|\lambda| < \rho$ is called the spectral radius of the operator $A$. Clearly $\rho(A) \leq \|A\|$ but we also have Gel'fand's formula:

$$
\rho(A) = \lim_{n \to \infty} \sqrt[n]{\|A^n\|} = \inf_n \sqrt[n]{\|A^n\|}.
$$

The existence of the limit and its equality to the infimum is guaranteed by the theorem of Fekete that if a sequence $(\alpha_n)$ of real numbers is subadditive, i.e.

$$
\alpha_{n+m} \leq \alpha_n + \alpha_m,
$$

then $\alpha_n$ tends to the infimum of the sequence as $n \to \infty$. Further, the outer radius of convergence $r$ of the series (23) is equal to the stated limit and $r \geq \rho(A)$ since the domain of convergence of the series is contained in $\text{reg } A$. On the other hand, $r \leq \rho(A)$ since the resolvent is analytic on the domain $|\lambda| > \rho(A)$.

If $\rho(A) = 0$, i.e. $\text{spec } A = \{0\}$, the operator $A$ is said to be quasinilpotent. In particular, any bounded nilpotent operator is of this type, however the class of quasinilpotent operators is more extensive than this.

**Example.** On $L^p(0, 1) (1 \leq p \leq \infty)$ the integration operator

$$
(I\phi)(s) = \int_{0}^{s} \phi(t) \, dt
$$

is not nilpotent but it is quasinilpotent. In fact

$$
(I^n\phi)(s) = \frac{1}{(n-1)!} \int_{0}^{s} (s-t)^{n-1} \phi(t) \, dt \quad (n = 1, 2, 3, \ldots),
$$

hence

$$
\|I^n\| = O\left(\frac{1}{n!}\right).
$$

Consequently, $\rho(I) = 0$ by Gel'fand's formula.

For any bounded operator $A$ defined on the whole space the circle $|\lambda| = \rho(A)$ contains points of the approximate spectrum.

Let $A$ be an isometric operator defined on a closed subspace $D(A) \subset E$. Its image $\text{Im } A$ is clearly also closed. Since

---

35 Hence it follows that the functional calculus for the operator $A$ operates on the disk $|\lambda| < \rho(A)$. 

all points $\lambda$ with $|\lambda| \neq 1$ are quasiregular. The defect number $n_\lambda$ is constant in the disk $|\lambda| < 1$ and in its exterior $|\lambda| > 1$. Its value on the disk is equal to $n_0 = \dim(E/\Im A)$. We denote its value on the exterior by $n_{\infty}$. It follows from the formula $\Delta_\lambda(A) = \Delta_{\lambda^{-1}}(A^{-1})$ ($|\lambda| \neq 1$) that $n_{\infty} = \dim(E/D(A))$ so that $n_{\infty} = 0 \iff D(A) = E$ (clearly $n_0 = 0 \iff \Im A = E$). The pair $(n_0, n_{\infty})$ is called the deficiency index of the isometric operator. If there exist quasiregular points in the unit disk, then $n_{\infty} = n_0$. In order that an isometric operator map the whole space onto itself (in a Hilbert space this means that it is unitary) it is necessary and sufficient that it have deficiency index $(0, 0)$.

The approximate spectrum of any isometric operator is unitary, i.e. it lies on the unit circle. If the operator is defined on the whole space then its spectrum lies in the disk $|\lambda| \leq 1$, the circle $|\lambda| = 1$ contains the points of the approximate spectrum but it can also contain quasiregular points; then they are regular and $n_0 = 0$. If $n_0 \neq 0$, the spectrum is the disk $|\lambda| \leq 1$ and the approximate spectrum is the circle $|\lambda| = 1$. The spectrum of an operator which maps the whole space isometrically onto itself lies on the unit circle.

To conclude this section let us consider the structure of the spectrum of a compact operator. This is the next in complexity after the algebraic case. All linear operators on a finite-dimensional LTS are compact and algebraic; the spectrum of any of these coincides with its set of eigenvalues.

Theorem. The spectrum of a compact operator $A$ on an infinite-dimensional LTS $E$ consists of zero and a no more than countable set of eigenvalues different from zero. The unique limit point of this set, if it is infinite, is zero.

Proof. Each point $\lambda \in \spec A \setminus \{0\}$ is an eigenvalue by the Fredholm alternative. The point $\lambda = 0$ belongs to the spectrum since otherwise the operator $A^{-1}A = 1$ is compact, which is impossible for an infinite-dimensional $E$.

Let $U$ be a balanced neighbourhood of zero whose image is relatively compact. Let us suppose that there exists a set $\{\lambda_n\}_{n=1}^\infty$ of eigenvalues such that $|\lambda_n| \geq \varepsilon > 0$; let $e_n (n = 1, 2, \ldots)$ be corresponding eigenvectors. Now consider $L_n = \Lin\{e_k\}_1^n$; clearly $\dim L_n = n$. We can choose in $L_n \cap \bar{U}$ a vector $x_n \notin L_{n-1} + U$. Since $Ax_n - \lambda_n x_n \in L_{n-1}$ and $Ax_m \in L_{n-1} (n > m)$ we have $Ax_n - Ax_m \notin \varepsilon U$, which is a contradiction. □

Let us enumerate the eigenvalues which are different from zero (if they exist): $\lambda_1, \lambda_2, \ldots$. We denote their orders by $v_1, v_2, \ldots$, and their maximal root subspaces by $W_1, W_2, \ldots$. They are all finite-dimensional and have invariant complements $V_i = \Im(A - \lambda_i 1)^{-1}$. Moreover $\spec(A|_{W_i}) = \{\lambda_k\}$ and $\spec(A|_{V_i}) = \spec(A \setminus \{\lambda_k\})$. Hence it follows that on a Banach space $E$ the resolvent $R_\lambda(A)$ has a pole of order $v_k$ at the point $\lambda_k$ and has no other singularities apart from the points $\lambda_k$ and zero, i.e. it is meromorphic on $\C \setminus \{0\}$.
Chapter 2. Foundations and Methods

If a compact operator has no eigenvalues different from zero (i.e. it is quasi-nilpotent in the Banach situation), it is said to be a Volterra operator. The following example explains the origin of this term.

**Example.** In $L^p(0, 1) (1 \leq p \leq \infty)$ Volterra's integral operator

$$(A\phi)(s) = \int_0^s R(s, t)\phi(t) \, dt \quad (0 \leq s \leq 1)$$

with continuous kernel $R(s, t)$ is a Volterra operator. In particular the integration operator $I$ is a Volterra operator.

Let $A$ be a non-negative compact operator on a Banach space $E$ with closed cone $K$. Let us suppose that the cone $K$ is total, i.e. $\overline{\text{Lin}} \, K = E$. If the spectral radius $\rho = \rho(A)$ is positive, then $\rho$ is an eigenvalue for $A$ and $A^*$ and corresponding to it there are an eigenvector $x \in K$ and an eigenfunctional $f \in K^*$ with $f(x) > 0$ (M.A. Rutman, 1938). The condition $\rho(A) > 0$ is satisfied if there exists a vector $v \geq 0$ ($v \neq 0$) such that $Av \geq cv$ for some $c > 0$ (then $\rho(A) \geq c$).

**Example.** Let us consider the integral operator

$$(A\phi)(s) = \int_0^{q(s)} R(s, t)\phi(t) \, dt \quad (0 \leq s \leq 1)$$

on $C[0, 1]$ ($R(s, t)$, $q(s)$ are continuous functions, $R(s, t) \geq 0$, $0 \leq q(s) \leq 1$). If there exists a point $s_0$ at which $R(s_0, s_0) > 0$ and $q(s_0) > s_0$ then $\rho(A) > 0$. In fact $A \geq 0$ and for $v(t)$ we can take the function which is equal to $\varepsilon - |t - s_0|$ for $|t - s_0| \leq \varepsilon$ and equal to zero outside of this interval (where $\varepsilon > 0$ is sufficiently small).

4.5. One-Parameter Semigroups. The representations $T(t)$ of the additive semigroup $\mathbb{R}_+$ of non-negative numbers with $T(0) = 1$ are designated by the term in the title. The representations of the additive group $\mathbb{R}$ are called one-parameter groups.

Each bounded linear operator $A$ on a Banach space $E$ generates a one-parameter group

$$T(t) = e^{At} = \sum_{n=0}^{\infty} \frac{A^n t^n}{n!} \quad (t \in \mathbb{R}).$$

Moreover

$$\frac{d}{dt} [T(t)x] = AT(t)x = T(t)Ax \quad (x \in E, t \in \mathbb{R}),$$

i.e. the vector-function $u(t) = T(t)x$ is the solution of the Cauchy problem

$$\dot{u}(t) = Au(t) \quad (-\infty < t < \infty), \quad u(0) = x$$

(the solution can easily be shown to be unique). The operator $A$ is infinitesimal for the representation $T$ in the sense that $A = T'(0)$. Because the operator $A$ is
bounded, the representation $T$ is continuous not only in the strong topology but also in the uniform topology:

$$\lim_{t \to 0} \|T(t + \tau) - T(t)\| = 0 \quad (t \in \mathbb{R}).$$

Conversely, if a representation $T(t) (t \in \mathbb{R}_+)$ is uniformly continuous, then $T(t) = e^{At}$, where $A$ is some bounded operator. This follows from the general theory which is presented below.

Let $T(t)$ be any one-parameter semigroup. Since $T(t + s) = T(t)T(s)$, the function $\alpha(t) = \ln \|T(t)\|$ is subadditive, $\alpha(t + s) \leq \alpha(t) + \alpha(s)$, and, by the continuous variant of Fekete's theorem, the limit

$$\sigma = \lim_{t \to \infty} \frac{\ln \|T(t)\|}{t} < \infty$$

exists. The Laplace transform

$$\tilde{T}(\lambda) = -\int_0^\infty T(t)e^{-\lambda t} dt$$

is therefore defined on the half-plane $\Pi_\sigma = \{\lambda: \text{re } \lambda > \sigma\}$. Since $\|T(t)\| \leq Me^{\sigma t}$ for some $c > \sigma, M > 0$, we have

$$\|\tilde{T}(\lambda)\| \leq \frac{M}{\text{re } \lambda - c} \quad (\lambda \in \Pi_\sigma).$$

The values of the operator function $\tilde{T}(\lambda)$ commute pairwise and satisfy the Hilbert identity

$$\tilde{T}(\lambda) - \tilde{T}(\mu) = (\lambda - \mu)\tilde{T}(\lambda)\tilde{T}(\mu).$$

The subspace $L = \text{Im} \tilde{T}(\lambda)$ therefore does not depend on $\lambda$. The operator $\tilde{T}(\lambda)$ maps $E$ bijectively onto $L$ (injectivity follows from the uniqueness theorem for the Laplace transform, since, if $\tilde{T}(\lambda)x = 0$, then $\tilde{T}(\mu)x = 0$ for all $\mu \in \Pi_\sigma$). Let us put $A(\lambda) = \lambda 1 + [\tilde{T}(\lambda)]^{-1}$. This does not depend on $\lambda$ and defines a closed operator $A (\mathcal{D}(A) = L)$ with $R_\lambda(A) = \tilde{T}(\lambda)$. The subspace $L$ is dense since

$$[\tilde{T}(\lambda)]^* = -\int_0^\infty T^*(t)e^{-\lambda t} dt$$

and, analogous to the previous situation, $\text{Ker}[\tilde{T}(\lambda)]^* = 0$.

We note that, since $R_\lambda(A)$ commutes with all $T(s)$, the operator $A$ also has this property.

Let us consider the Cauchy problem

$$\dot{u}(t) = Au(t) \quad (t \geq 0), \quad u(0) = x \in \mathcal{D}(A).$$

If $\mu \in \Pi_\sigma$ and $y = Ax - \mu x$, then

$$x = R_\mu y = -\int_0^\infty (T(s)y)e^{\mu s} ds,$$
and hence
\[
T(t)x = - \int_0^\infty (T(t + s)y)e^{-\lambda s}ds = -e^{\mu t} \int_t^\infty (T(s)y)e^{-\lambda s}ds.
\]
Consequently the function \(T(t)x\) has a continuous derivative for \(t \geq 0\) and
\[
\frac{d}{dt} [T(t)x] = \mu T(t)x + T(t)y = T(t)Ax = AT(t)x \quad (t \geq 0),
\]
i.e. \(T(t)x\) is a solution of the Cauchy problem (25). In particular, putting \(t = 0\)
we obtain \(Ax = \frac{d}{dt} [T(t)x]_{t=0}\).

The Cauchy problem (25) can be considered\(^{36}\) for any operator \(A\). In the
situation described above it does not have solutions other than \(T(t)x\), i.e. the
corresponding homogeneous problem \(\dot{v}(t) = Av(t)\) \((t \geq 0)\), \(v(0) = 0\) has only the
trivial solution. This is the case (Yu.I. Lyubich, 1960) even under the much weaker
restriction of growth of the resolvent on the ray \(\lambda > \lambda_0 > 0\):
\[
\lim_{\lambda \to +\infty} \frac{\ln \|R_\lambda\|}{\lambda} < \infty
\]
(this cannot be relaxed further). In fact
\[
R_\lambda v(0) = -\int_0^s v(t)e^{\lambda s\lambda}dt + e^{-\lambda s}R_\lambda v(s) \quad (s > 0),
\]
from which it is obvious that \(R_\lambda v(0)\) is the local Laplace transform for \([-v(t)]\).
If \(v(0) = 0\) then \(v(t) \equiv 0\) by the uniqueness theorem for the l.L.t.

The subspace \(I(A) \subset D(A)\) of those \(x\) for which the Cauchy problem has a
solution (independently of uniqueness\(^{37}\)) is called the initial subspace. It can fail
to be dense (and can actually be the zero subspace) even in the case where \(D(A)\)
is dense. But if a bound of the form \(\|R_\lambda\| = O(|\lambda|^m)\) is satisfied in some half-
plane \(\Pi_c\) contained in reg \(A\) and \(D(A)\) is dense, then \(I(A)\) is dense (Yu.I. Lyubich,
1964).

If \(x \in I(A)\) and \(u(t)\) is a corresponding solution, then clearly its values belong
to \(I(A)\) for all \(t\) (in this sense \(I(A)\) is invariant). If the Cauchy problem is
determinate, the evolutionary operator \(U(t)\) transforming \(u(0)\) into \(u(t)\) \((t \geq 0)\) is
defined on \(I(A)\). Clearly \(U(0) = 1\) and \(U(t + s) = U(t)U(s)\). If \(U(t)\) is bounded for
each\(^{38}\) \(t > 0\), we have a representation of the semigroup \(R_\lambda\) in the subspace \(I(A)\)
(in general it is not closed). However if \(\sup_{0 < t \leq \varepsilon} \|U(t)\| < \infty\) for some \(\varepsilon > 0\),

\(^{36}\) It is implicit that the solution is continuously differentiable and its values belong to \(D(A)\) for all
\(t \geq 0\) (in a relaxed formulation – for \(t > 0\)).

\(^{37}\) A Cauchy problem for which the solution is unique is said to be determinate (this property does
not depend on \(x\)).

\(^{38}\) In this case the Cauchy problem is said to be pointwise stable on \(I(A)\) (simply pointwise stable if
\(I(A)\) is dense).
then the extension by continuity of $U(t)$ onto $I(A)$ preserves continuity of the trajectories and we obtain a representation of the semigroup $R_+$ in the closed subspace $I(A)$. Its infinitesimal operator is the closure of the operator $A$ (S.G. Krejn – P.E. Sobolevskij, 1958).

The infinitesimal operator of a one-parameter semigroup is also said to be generating for this semigroup or to be a generator of it since, clearly, the semigroup with a given infinitesimal operator is unique. The question arises of describing the class of generators of all possible one-parameter semigroups.

**Miyadera-Feller-Phillips Theorem.** In order that a densely defined linear operator $A$ on a Banach space be the infinitesimal operator of a one-parameter semigroup, it is necessary and sufficient that $\text{reg } A$ contain a half-plane $\Pi_c$ in which the inequality

$$\|R^n_\lambda\| \leq \frac{M}{(\Re \lambda - c)^n} \quad (n = 0, 1, 2, \ldots; M \text{ a positive constant})$$

is satisfied.

**Necessity.** This follows from the bound $\|T(t)\| \leq Me^{\alpha t}$ and the formula

$$R^n_\lambda = \frac{1}{(n-1)!} \frac{d^n}{d\lambda^{n-1}} \left[ \frac{d}{d\lambda} R_\lambda \right] = (-1)^n \frac{1}{(n-1)!} \int_0^\infty T(t)t^n e^{-\lambda t} dt.$$ 

**Sufficiency.** Under the given conditions the Cauchy problem (25) is determinate and the initial subspace $I(A)$ is dense. Since

$$\int_0^s u(t)e^{-\lambda t} dt = R_\lambda(u(0)e^{-\lambda t} - u(0)),$$

we have by Widder's inversion formula ($0 < t < s$):

$$u(t) = \lim_{n \to \infty} \left\{ (-\lambda R^n_\lambda u(0) + \lambda^n \sum_{k=0}^n \frac{(1)}{(n-k)!} \frac{d^k}{d\lambda^k} R_\lambda u(s) \right\}_{\lambda = n/t}.$$ 

Hence it follows that $\|u(t)\| \leq Me^{\alpha t} \|u(0)\|$, i.e. for the evolutionary operator: $\|U(t)\| \leq Me^{\alpha t}$. Since $A$ is closed under the conditions of the theorem, it is the infinitesimal operator of the representation $U$. □

An immediate consequence of this result (taking account of the connection between the bounds for the representation and the resolvent) is the

**Hille-Yosida Theorem.** In order that a densely defined linear operator $A$ on a Banach space be the infinitesimal operator of a one-parameter semigroup of contractions, it is necessary and sufficient that the right half-plane $\Pi_0$ be contained in $\text{reg } A$ and that the inequality

In this case the Cauchy problem is said to be stable on $I(A)$ (simply stable if $I(A)$ is dense).
be satisfied in $\Pi_0$.

The operator $A$ in this case is said to be \textit{dissipative}.

\textbf{Example.} The semigroup $\mathbb{R}_+$ acts by means of shifts on $L^p(\mathbb{R}^+)$: $(T(t)\phi)(s) = \phi(s + t) \ (s, t \geq 0)$. Clearly $\|T(t)\| \leq 1$. The infinitesimal operator of this regular representation is equal to $\frac{d}{ds}$. It is defined on the absolutely continuous functions which, together with the derivative, belong to $L^p$. This is a dissipative operator.

The theory of one-parameter groups is reduced to the previous case by considering representations of the semigroups $\mathbb{R}_+$. In particular, the infinitesimal operators of one-parameter groups of isometries can be characterised (I.M. Gel'fand, 1939) as those operators whose spectrum lies on the imaginary axis and for which the bound

$$\|R_\lambda\| \leq \frac{1}{\text{re } \lambda} \quad (\text{re } \lambda \neq 0)$$

is satisfied. These operators are said to be \textit{conservative}.

\textbf{Example.} A regular representation of the group $\mathbb{R}$ in $L^p(\mathbb{R})$ is isometric. Its infinitesimal operator $\frac{d}{ds}$ can be described just as in the case of $\mathbb{R}_+$, but now it is actually conservative.

Conservative and dissipative operators have many remarkable properties, in particular the spectral radius of a bounded conservative operator $A$ on a Banach space $E$ is equal to its norm (V.Eh. Katselelson, 1970). In fact, the function $f(e^{-x}) \ (x \in E, f \in E^*)$ is an entire function of exponential type $\leq \rho(A)$ and it is bounded on the real axis. From Bernstein's inequality we have

$$|f(e^{it}x)| = \left|\frac{d}{dt} (f(e^{it}x))\right| \leq \rho(A) \sup |f(e^{it}x)| \leq \rho(A) \|f\| \cdot \|x\|.$$ 

Putting $t = 0$ we obtain $|f(Ax)| \leq \rho(A) \|f\| \cdot \|x\|$ and consequently $\|A\| \leq \rho(A)$.

Incidentally, Bernstein's inequality itself can be interpreted as the coincidence of the spectral radius with the norm of the differentiation operator on the space $B_\sigma$ of entire functions of exponential type $\leq \sigma$ which are bounded on the real axis (it is remarkable that the differentiation operator on $B_\sigma$ is bounded).

A general mechanism underlying inequalities of Bernstein type and results pertaining to it in spectral theory was brought to light by E.A. Gorin (1980). It was based on the functional calculus associated with integral representations of positive functions.

From the point of view of the Cauchy problem there is great interest in the study of asymptotic (as $t \to \infty$) properties of one-parameter semigroups. For example, boundedness (i.e. $\sup \|T(t)\| < \infty$) is equivalent to Lyapunov stability.
for the corresponding Cauchy problem. For bounded one-parameter semigroups on a reflexive Banach space we have the ergodic theorem: the strong limit

$$\lim_{s \to \infty} \frac{1}{s} \int_0^s T(t) \, dt$$

exists. It is a bounded projection (for semigroups of contractions it is an orthogonal projection) on the subspace of fixed points \( \{ x: T(t)x = x \, (t \geq 0) \} = \ker A \) (\( A \) is the infinitesimal operator). This theorem owes its origin to the analysis of the foundations of statistical mechanics, which goes back to Koopman and von Neumann\(^{40}\).

There is an important class of representations for which the ergodic theorem is true without restriction on the space – almost-periodic representations.

A representation \( T \) is said to be almost-periodic (a.p.) if all of its orbits are relatively compact\(^{41}\).

**Example 1.** In the space \( AP(\mathbb{R}) \) of almost-periodic functions on \( \mathbb{R} \) the representation of the group \( \mathbb{R} \) by means of shifts (the regular representation) is a.p.

The definition of an a.p.f. carried over without change to functions on the semiaxis \( \mathbb{R}_+ \). We have correspondingly a regular a.p. representation of the semigroup \( \mathbb{R}_+ \) in \( AP(\mathbb{R}_+) \).

The general form of an a.p.f. on the semiaxis\(^{42}\) is \( \phi = \phi_0 + \phi_1 \), where \( \phi_1(t) \) tends to zero as \( t \to \infty \), and \( \phi_1 \) can be extended to the whole axis as an a.p.f. (Fréchet, 1941).

**Example 2.** Let \( A \) be a compact operator on a Banach space \( E \) which does not contain any subspaces isomorphic to the space \( c \) of convergent sequences \( (\xi_n)_{n=0}^\infty \) (in particular, \( E \) may be reflexive). Then, if the representation \( e^{st} (t \in \mathbb{R}) \) is bounded, it is a.p. In fact, for the function \( u(t) = e^{xt} (x \in E) \) the derivative \( \dot{u}(t) = Au(t) \) is a.p. because of the compactness of \( A \) and the function itself is bounded. Consequently \( u(t) \) is a.p. (here we are using a generalisation due to M.I. Kadets (1969) of the Bohl-Bohr theorem on the integral of an a.p.f.).

By restricting an a.p. representation of the group \( \mathbb{R} \) to the semigroup \( \mathbb{R}_+ \) we obtain, clearly, an a.p. representation but it is such that all the operators \( T(t) \) \( (t \in \mathbb{R}_+) \) are invertible and their inverses are bounded (in this sense it is a group representation of the semigroup \( \mathbb{R}_+ \)). There is a different kind of a.p. representation of the semigroup \( \mathbb{R}_+ \), all the orbits of which tend to zero as \( t \to \infty \). Further, if the space \( E \) can be decomposed into the topological direct sum of the spaces

\(^{40}\)This author established (1932) the ergodic theorem in Hilbert space. The generalisation given here is due to Loreh. Usually formulations of ergodic theorems are given for semigroups of powers \( (\mathbb{T})^n \). The 'continuous' case considered in the text is easily reduced to the 'discrete' case (such a reduction occurs quite often). Further, as a rule, we restrict ourselves to only one of these cases.

\(^{41}\)In exactly the same way a.p. is defined for bounded representations of any topological semigroups in any LTS. The property of boundedness is not required a priori in a barrelled LCS since it follows from the relative compactness of the orbits by the Banach-Steinhaus theorem.

\(^{42}\)In general we can define a.p.f.s on any topological semigroup; they can be not only scalar-valued but also vector-valued.
$E_1, E_0$ in which a.p. representations $T_1, T_0$ are given (for example, $T_1$ a group representation and $T_0$ having orbits tending to zero) then the natural representation $T_1 + T_0$ defined on $E$ is a.p. We have the boundary spectrum splitting-off theorem (M.Yu. Lyubich – Yu.I. Lyubich, 1984): let $T$ be an a.p. representation of the semigroup $\mathbb{R}_+$ in a Banach space $E$; then $E$ can be decomposed into the topological direct sum of invariant subspaces $E_1, E_0$ such that $T_1 = T|_{E_1}$ is a group representation and the orbits of $T_0 = T|_{E_0}$ tend to zero. The subspaces $E_1, E_0$ are called the boundary and interior subspaces respectively and the projection $P$ associated with them ($\text{Im} \ P = E_1, \text{Ker} \ P = E_0$) is called the boundary projection. It arises in the following way. Corresponding to an a.p. representation $T$ there is a Bohr compact set – the strong closure of the family (semigroup) of operators $\{T(t)\}_{t \geq 0}$. This is an Abelian semigroup which is compact in the strong topology. In any compact semigroup there is a smallest two-sided ideal – the Sushkevich kernel (A.K. Sushkevich, 1928; Numakura, 1952). For an Abelian semigroup (and in certain other cases) the Sushkevich kernel is a compact group. The boundary projection for an a.p. representation is the identity of the Sushkevich kernel of the Bohr compact set. Hence it is clear, incidentally, that $\|P\| \leq \sup \|T(t)\|$ and in particular, if $T$ is a contraction then $P$ is an orthoprojection. In the latter case the representation $T_1$ is an isometry. The theory which has been described carries over to a.p. representations of any Abelian (and certain non-Abelian) topological semigroups, in particular, to semigroups of powers of an a.p. operator. A prototype of the general theory of a.p. representations, the boundary spectrum splitting-off theorem for bounded semigroups of operators on a reflexive Banach space, was constructed by Jacobs (1957), after which De Leeuw and Glicksberg (1961) obtained corresponding results in the weak topology for any Banach space. In the weak a.p. case the orbits for $T_0$ do not tend to zero in general, even weakly, and they only have zero as an adherent point. Furthermore $T_0$ can turn out to be a group representation (for example, $(T(t)f)(s) = e^{itf}(s, t \geq 0)$ in $L^2(\mathbb{R}^+)$).

4.6. Conjugation and Closure. Let $A$ be a linear operator on an LTS $E$ and suppose that its domain of definition $D(A)$ is dense. Then the conjugate operator $A^*$ on $E^*$ can be defined by means of the relation $(A^*f)(x) = f(Ax)$ for those $f \in E^*$ for which the right hand side of the identity is continuous on $D(A)$ and, consequently, can be uniquely extended to a continuous linear functional on $E$. If $E$ is a Hilbert space the definition of the conjugate operator is usually modified on the basis of Riesz’s theorem so that $A^*$ acts on the space $E$ itself:

$$(Ax, y) = (x, A^*y) \quad (x \in D(A), y \in D(A^*))$$

Therefore $(\alpha A)^* = \overline{\alpha}A^*$. In this case we call $A^*$ the adjoint of $A$. 

---

43 The fundamental work of A.K. Sushkevich is concerned with finite semigroups.

44 In particular, any operator of the form $A = U + V$, where $U$ is compact, $\|V^*\| < 1$ for some $m$ and $\sup_{n \geq 0} \|A^n\| < \infty$ is of this type (Yosida – Kakutani, 1941).
Theorem. If a densely defined operator $A$ on an LTS $E$ has continuous inverse $A^{-1}: E \to D(A)$ then $A^*$ has inverse $(A^*)^{-1}: E^* \to D(A^*)$ and $(A^*)^{-1} = (A^{-1})^*$.

Thus, for example, $(A^*)^{-1}$ is continuous under the topology of uniform convergence on bounded sets and also in the $w^*$-topology.

Corollary. If $A$ is a densely defined operator on a Banach space, then $\text{reg } A^* = \text{reg } A$ and consequently $\text{spec } A^* = \text{spec } A$.

In the Hilbert space context $\text{reg } A^* = \overline{\text{reg } A}$ and $\text{spec } A^* = \overline{\text{spec } A}$ (the bar denotes complex conjugation). The defect space $E/\Delta_\lambda (\lambda \in \text{spec } A)$ is isomorphic to the subspace $\Delta_\lambda^\perp = N_\lambda \equiv \text{Ker}(A^* - \lambda I)$.

The preceding theorem has a converse in a Banach space: if $A$ is densely defined and closed and $A^*$ has a bounded inverse then $A$ also has a bounded inverse (consequently, $(A^{-1})^* = (A^*)^{-1}$).

An operator $A$ on a Hilbert space is said to be selfadjoint if $A^* = A$. For example, any orthoprojection is of this type. If $A \subseteq A^*$ which, in the case of dense $D(A)$, is equivalent to the identity

$$
(Ax, y) = (x, Ay) \quad (x, y \in D(A)),
$$

then $A$ is said to be a symmetric operator. Selfadjoint operators play a central role in quantum mechanics, where they are used to give a suitable description of physical quantities (coordinates, impulses, energy, etc.). The principal credit for constructing the spectral theory of selfadjoint operators is due to Hilbert and von Neumann.

We say that an operator $A$ on an LTS $E$ admits closure if the closure of its graph is the graph of some operator $\overline{A}$, the closure of the operator $A$. Clearly $A \subseteq \overline{A}$ and $\overline{A}$ is closed. In order that $A$ admit closure, it is necessary and sufficient that a closed extension exist for it; moreover $\overline{A}$ is the least closed extension for $A$.

A continuous operator always admits closure; this is just its extension by continuity to $\overline{D(A)}$. The domain of definition of a closed continuous operator is therefore closed and if it is also dense then it coincides with the whole space. Conversely, any continuous operator defined on a closed subspace is closed.

A conjugate operator $A^*$ is always closed (even $w^*$-closed). In particular, in a Hilbert space $E$ any selfadjoint operator is closed and any densely defined symmetric operator admits closure since $A \subseteq A^*$. Moreover it is clear that $\overline{A}$ is symmetric. If a symmetric operator $A$ is defined on the whole space $E$, then $A$ is a bounded selfadjoint operator (Hellinger-Toeplitz theorem). In fact, since $A \subseteq A^*$ and $D(A) = E$ we have $A = A^*$. But $A$ is then closed and consequently it is bounded by the closed graph theorem.

Theorem. If a densely defined operator $A$ on a Hilbert space admits closure, then $A^*$ is densely defined and $A^{**} = \overline{A}$.

A closed densely defined operator on a Hilbert space which commutes with its adjoint is said to be normal. In particular all unitary operators $(A^*A = AA^* = 1)$ are of this type.
4.7. Spectra and Extensions of Symmetric Operators. Let us consider a densely defined closed symmetric operator \( A \) on a Hilbert space \( E \). It follows from the identity (27) that
\[
\|Ax - \lambda x\| \geq |\text{im } \lambda| \|x\| \quad (x \in D(A)),
\]
from which it is obvious that all \( \lambda \) with \( \text{im } \lambda \neq 0 \) are quasiregular. The defect numbers \( n_+, n_- \) associated with the half-planes \( \text{im } \lambda > 0, \text{im } \lambda < 0 \) are in general unequal, however, if there is a real quasiregular point, then \( n_+ = n_- \). The pair \( (n_+, n_-) \) is called the deficiency index of the operator \( A \).

**Theorem 1.** In order that \( A \) be selfadjoint, it is necessary and sufficient that its deficiency index be \((0, 0)\), i.e. that \( \text{spec } A \subset \mathbb{R} \).

**Necessity.** If \( A^* = A \) then \( N_\xi = 0 \) whenever \( \text{im } \lambda \neq 0 \), since all eigenvalues of a symmetric operator are real.

**Sufficiency.** If \( n_+ = n_- = 0 \) there exists for any \( \lambda \) (\( \text{im } \lambda \neq 0 \)) and any \( x \in D(A^*) \) a \( y \in D(A) \) such that \( Ay - \lambda y = A^*x - \lambda x \) and hence \( A^*(x - y) = \lambda(x - y) \). Consequently \( x - y = 0 \), i.e. \( x = y \in D(A) \).

If \( \tilde{A} \) is a closed symmetric extension of the operator \( A \) then \( \tilde{n}_+ \leq n_+ \) and \( \tilde{n}_- \leq n_- \). The operator \( \tilde{A} \) is 'closer' to selfadjointness than \( A \) since\(^45\) \( A \subset \tilde{A} \subset \tilde{A}^* \subset A^* \). Is it possible to choose \( \tilde{A} \) to be selfadjoint, i.e. does \( A \) admit a selfadjoint extension? This important question was investigated by von Neumann (1929) with the help of the Cayley transform which is defined for an operator \( A \) and any \( \lambda \) (\( \text{im } \lambda \neq 0 \)) by the formula
\[
V_\lambda = V(A)(\lambda - 1)/V(A - \lambda I)^{-1}.
\]
The homomorphism \( V_\lambda : \Delta_\lambda \to \Delta_\lambda \) is surjective and isometric\(^46\) since if \( \lambda = \alpha + i\beta \) we have
\[
\|Ax - \lambda x\|^2 = \|Ax - \alpha x\|^2 + \beta^2 \|x\|^2 \quad (x \in D(A))
\]
and hence \( \|Ax - \lambda x\| = \|Ax - \lambda x\| \). Moreover the subspace \( \text{Im}(V_\lambda - I) = D(A) \) is dense. Conversely, any isometric operator \( V \) for which \( \text{Im}(V - I) \) is dense is the Cayley transform of the densely defined closed symmetric operator \( A = (\lambda V - \lambda I)(V - I)^{-1} \) (it follows from the denseness of \( \text{Im}(V - I) \) that unity is not an eigenvalue of the operator \( V \) and therefore \( (V - I)^{-1} : \text{Im}(V - I) \to D(V) \) is defined).

By Theorem 1 the Cayley transform of a selfadjoint operator (and only of such an operator) is unitary. The question about selfadjoint extensions therefore reduces to the existence of unitary extensions of the isometric operator \( V_\lambda \), which in turn is equivalent to the defect subspaces \( N_\lambda, N_\lambda^* \) being isometric. Thus we have

\(^45\) For any two linear operators \( A_1, A_2 \) with dense domains of definition the inclusion \( A_1 \subset A_2 \) implies that \( A_1^* \supset A_2^* \).
\(^46\) Implicit in the construction of the Cayley transform is the idea of spectral mapping: the linear-fractional transformation \( \zeta \mapsto (\zeta - \lambda)(\zeta - \bar{\lambda})^{-1} \) maps the real axis onto the unit circle, therefore its operator analogue must transform symmetric operators into isometric operators.
Theorem 2. In order that $A$ admit a selfadjoint extension, it is necessary and sufficient that its defect numbers be equal.

From what has been said it is also evident that in the case $n_+ = n_-$ the set of selfadjoint extensions can be parameterised by the isometries of $N_\lambda$ onto $N_\lambda$, in particular, for differential operators (I.M. Glazman, 1949).

Example 1. Let us consider the operator $D = \frac{1}{i} \frac{d}{ds}$ on $L^2(0, 1)$ which is defined for all absolutely continuous functions with derivative in $L^2$. We will denote by $D_0$ its restriction under the boundary conditions $\phi(0) = \phi(1) = 0$. This is a closed symmetric operator and $D_0 = D$. Each $\lambda \in \mathbb{C}$ is an eigenvalue for $D$, the eigenspace $N_\lambda$ is one-dimensional and it is generated by the function $e_\lambda(s) = e^{i\lambda s}$. The deficiency index of the operator $D_0$ is therefore equal to $(1, 1)$. The isometries of $N_\lambda \rightarrow N_\lambda$ take the form $V_\lambda(\theta)e_\lambda = \theta e^{-\beta} e_\lambda$ ($\beta = \text{im } \lambda, \theta$ a parameter, $|\theta| = 1$). The selfadjoint extension $D^\theta$ of the operator $D_0$ corresponding to the isometry $V_\lambda(\theta)$ is such that

$$(D^\theta - \lambda I)(D^\theta - \overline{\lambda} I)^{-1} e_\lambda = \theta e^{-\beta} e_\lambda.$$ 

Putting $(D^\theta - \lambda I)^{-1} e_\lambda = v$, we obtain

$$D^\theta v - \lambda v = e_\lambda, \quad D^\theta v - \lambda v = \theta e^{-\beta} e_\lambda,$$

and hence

$$v = \frac{e_\lambda - \theta e^{-\beta} e_\lambda}{\lambda - \overline{\lambda}}.$$ 

Set $\lambda = i$. Then $\nu(1) = \omega v(0)$, where $\omega = (1 - e\theta)(e - \theta)^{-1}$ is a parameter which, together with $\theta$, runs through the unit circle. Thus the extension $D^\theta$ of the 'minimal' operator $D_0$ can be described as the restriction of the 'maximal' operator $D$ under the boundary conditions $\phi(1) = \omega \phi(0)$.

Example 2. The operator $D$ on $L^2(0, \infty)$ defined as in the previous example can be restricted to $D_0$ by imposing the condition $\phi(0) = 0$. As before, $D_0^* = D$ but now the deficiency index is $(0, 1)$ since $e_\lambda \in L^2(0, \infty)$ only if $\text{im } \lambda > 0$. Consequently $D_0$ does not have selfadjoint extensions (or even symmetric extensions other than itself).

Example 3. The operator $D$ on $L^2(-\infty, \infty)$ defined as before is selfadjoint.

The operator $iD$ on $L^2(-\infty, \infty)$ is conservative and only operators of the form $iA (A = A^*)$ are conservative on a Hilbert space (see inequality (28)). Therefore if $A$ is a bounded selfadjoint operator we have $\rho(A) = \|A\|$ (which can also be established by elementary means). From this it is easy to establish the non-emptiness of the spectrum of any selfadjoint operator $B$. In fact if $0 \notin \text{spec } B$ then $A = B^{-1}$ is a bounded selfadjoint operator and $A \neq 0$. Therefore $\rho(A) = \|A\| > 0$. Now at least one of the numbers $\pm \rho(A)$ is in $\text{spec } A$ and then its reciprocal must be in $\text{spec } B$. We note incidentally that $\rho(A) = 0 \Rightarrow A = 0$ (i.e. a non-zero self-
adjoint operator cannot be quasinilpotent). Even if the selfadjoint operator $A$ is not presupposed to be bounded but $\text{spec } A = \{0\}$ then $A = 0$. In fact, in this case $R_A(\lambda \in \mathbb{R}, \lambda \neq 0)$ is a bounded selfadjoint operator and $\text{spec } R_A = \{-\lambda^{-1}\}$. Consequently $\|R_A\| = |\lambda|^{-1}$ which leads to the inequality $\|Ax\|^2 - 2\langle Ax, x \rangle \geq 0$ for all $x \in D(A)$ and $\lambda \in \mathbb{R} \setminus \{0\}$. But then $(Ax, x) \equiv 0$ and hence $A = 0$.

The last conclusion can be made on the basis of the polarisation formula

$$(Ax, y) = \frac{1}{4} \sum_{k=0}^{3} i^k (A(x + i^ky), x + i^ky),$$

which is valid for any linear operator $A$ on a Hilbert space $E$. Incidentially, it follows from this formula that if a Hermitian quadratic functional $(Az, z)$ $(z \in D(A))$ is bounded on the unit sphere then the operator $A$ is also bounded and $\|A\| \leq 2 \sup_{\|x\|=1} |(Az, z)|$ (the converse is obvious).

The functional $(Ax, y)$ $(x \in D(A), y \in E)$ is sesquilinear, i.e. linear in $x$ and semilinear in $y$, i.e. it is additive and if $y$ is replaced by $xy$ then it is multiplied by $\bar{x}$. For any sesquilinear functional $\beta(x, y)$ which is defined for $x$ in some subspace $L$ and for all $y \in E$ there exists by Riesz's theorem a unique linear operator $A$ such that $\beta(x, y) = (Ax, y)$ $(x \in D(A) = L, y \in E)$. Symmetry of the operator $A$ is equivalent to Hermitian symmetry of the functional $\beta(y, x) = \beta(x, y)$, which, in turn, is equivalent to the quadratic functional $\beta(x, x)$ being real.

The operator $iD$ on $L^2(0, \infty)$ is dissipative and in general all operators on a Hilbert space of the form $iA$ $(A \in A^*, n_+ = 0)$ and their adjoints are dissipative but these are by no means all the dissipative operators (for example, $(-I)$, where $I$ is the integration operator on $L^2(0, 1)$). The following criterion for dissipativity of a closed densely defined operator $A$ on a Hilbert space is an easy consequence of the Hille-Yosida theorem:

$$\Re(Ax, x) < 0 \quad (x \in D(A)); \quad \Re(A^*y, y) \leq 0 \quad (y \in D(A^*)).$$

**Remark.** Dissipative operators on a Hilbert space are often defined only by the requirement $\Re(Ax, x) < 0$ (which in addition they rotate through $\pi/2$. $\Im(Ax, x) \geq 0$). Under this the half-plane $\Re \lambda > 0$ is quasiregular but we can have $n_+ > 0$. For a dissipative operator in our sense $n_+ = 0$. We note in this connection that symmetric operators with zero defect number and only such operators are maximal in the sense that they do not admit symmetric extensions (since their Cayley transforms do not admit isometric extensions).

A symmetric operator $A$ is said to be non-negative $(A \geq 0)$ if $(Ax, x) \geq 0$, positive $(A > 0)$ if $(Ax, x) > 0 \quad (x \neq 0)$ and positive definite $(A \gg 0)$ if $(Ax, x) \geq c\|x\|^2$ $(c$ is a positive constant) (in all cases $x \in D(A)$). For example, any orthoprojection is non-negative, multiplication by $c > 0$ is a positive definite operator. If $\langle e_n \rangle_0^\infty$ is an orthonormal basis and $\langle \lambda_n \rangle_0^\infty$ is a sequence of positive numbers which tends to zero, the corresponding diagonal operator

\[^4^7\] Dissipativity is preserved under conjugation even in the Banach space case.
Ax = \sum_{n=0}^{\infty} \lambda_n \xi_n e_n \quad \left( x = \sum_{n=0}^{\infty} \xi_n e_n \right)

is positive but it is not positive definite.

For a non-negative operator \( A \) all points \( \lambda < 0 \) are quasiregular and therefore \( n_+ = n_- \). If \( A^* = A \geq 0 \), the spectrum of the operator \( A \) is non-negative. The spectrum of a positive definite selfadjoint operator lies on the corresponding semiaxis \( \lambda \geq c \).

In the real Banach space of bounded selfadjoint operators the non-negative operators form a closed solid cone (1 is an interior point) as a result of which we can work with inequalities in this space. This approach improves substantially the norm bounds.

### 4.8. Spectral Theory of Selfadjoint Operators

The simplest spectral theorem (after the algebraic case) concerns a compact selfadjoint operator \( A \) on a Hilbert space \( E \) and asserts that \( E \) coincides with the closure of the orthogonal sum of the eigenspaces \( E_n \) corresponding to all possible eigenvalues \( \{ \lambda_n \}^\infty_0 \) (zero is included if it is an eigenvalue). This theorem is proved just as in the finite-dimensional case. With a view to generalisation it is convenient to express it in the form of a spectral resolution

\[
Ax = \sum_{n=0}^{\infty} \lambda_n P_n x,
\]

where \( P_n \) is an orthoprojection onto \( E_n \) and

\[
x = \sum_{n=0}^{\infty} P_n x.
\]

A family of orthoprojections \( E_\lambda (-\infty < \lambda < \infty) \) on a Hilbert space \( E \) is called an orthogonal resolution of the identity if

1) \( E_\lambda E_\mu = E_\lambda \) \((\lambda < \mu)\),
2) \( \lim_{\lambda \to +\infty} E_\lambda x = x, \lim_{\lambda \to -\infty} E_\lambda x = 0 \) \((x \in E)\),
3) the function \( E_\lambda x \) \((x \in E)\) is continuous on the left with respect to \( \lambda \) on the whole axis.

Under these conditions for any interval \( \Delta = [a, b] \) the operator \( E_\Delta = E_b - E_a \) is an orthoprojection.

Any orthogonal resolution of the identity \( E_\lambda \) defines a family of measures \( m_x \) on \( \mathbb{R} \) which are generated by the non-decreasing functions \( (E_\lambda x, x) \) \((x \in E)\). Suppose that the scalar function \( \phi \) is measurable with respect to all the \( m_x \). It is then said to be measurable with respect to \( E_\lambda \). We define the integral.

---

48 But it does nevertheless include the Hilbert-Schmidt theorem on integral equations with Hermitian-symmetric kernels.

49 To be precise we assume that the set of eigenvalues is infinite. A compact selfadjoint operator with finite spectrum is algebraic.

50 An orthogonal resolution of the identity can be given on some interval, for example \([0, 2\pi]\), rather than on the whole axis. In this example the conditions analogous to 2) have the form \( E_0 - 0, E_{2\pi} - 1 \).
with respect to the operator-valued measure $\int d\xi_\lambda$ as the linear operator for which the subspace

$$D(\Phi) = \left\{ x : x \in E, \int_{-\infty}^{\infty} |\phi(\lambda)|^2 \, d(\xi_\lambda x, x) < \infty \right\}$$

(assumed further to be dense) serves as domain of definition and the identity

$$(\Phi x, y) = \int_{-\infty}^{\infty} \phi(\lambda) \, d(\xi_\lambda x, y) \quad (x, y \in D(\Phi))$$

is satisfied.

It can be shown that in addition

$$\Phi^* = \int_{-\infty}^{\infty} \overline{\phi(\lambda)} \, d\xi_\lambda.$$

Consequently it is necessary and sufficient for selfadjointness of the operator $\Phi$ that the function $\phi$ be real-valued almost everywhere (with respect to each measure $m_\lambda$). For boundedness of the operator $\Phi$ it is necessary and sufficient that the function $\phi$ be bounded in the sense that

$$\|\phi\| = \sup_x \left( \text{ess.sup}_{m_\lambda} |\phi| \right) < \infty$$

(in which case $\|\Phi\| = \|\phi\|$). In general

$$\|\Phi x\|^2 = \int_{-\infty}^{\infty} |\phi(\lambda)|^2 \, d(\xi_\lambda x, x) \quad (x \in D(\Phi)).$$

The construction which has been described is called the functional calculus for the given orthogonal resolution of the identity $\xi_\lambda$. The correspondence $\phi \mapsto \Phi$ is linear and $\phi_1, \phi_2 \mapsto \Phi_1 \Phi_2$ if at least one of the functions $\phi_1, \phi_2$ is bounded. The functional calculus is an isometric isomorphism on the algebra of bounded functions into the algebra of bounded linear operators.

**Example.** Let $A \subset \mathbb{R}$ be a Borel set and let $\chi_M$ be its characteristic function. Since $\chi_M^2 = \chi_M$ and $\bar{\chi}_M = \chi_M$, the operator

$$\xi_M = \int_{-\infty}^{\infty} \chi_M(\lambda) \, d\xi_\lambda = \int_M d\xi_\lambda$$

is an orthoprojection. In particular

$$\xi_R = \int_{-\infty}^{\infty} d\xi_\lambda = 1.$$

The spectral resolution of a compact selfadjoint operator $A$ can be expressed in the form
A = \int_{-\infty}^{\infty} \lambda \, d\mathcal{E}_\lambda,

where

\mathcal{E}_\lambda x = \sum_{\lambda_n < \lambda} P_n x \quad (x \in \mathcal{E}).

**Theorem.** For any selfadjoint operator $A$ there exists a (unique) orthogonal resolution of the identity $\mathcal{E}_\lambda (-\infty < \lambda < \infty)$ such that the spectral resolution

$$A = \int_{-\infty}^{\infty} \lambda \, d\mathcal{E}_\lambda$$

holds. Moreover

$$\phi(A) \overset{\text{def}}{=} \int_{-\infty}^{\infty} \phi(\lambda) \, d\mathcal{E}_\lambda$$

for all admissible\(^{51}\) $\phi$.

In particular we have the formula for the resolvent

$$R_\mu = \int_{-\infty}^{\infty} \frac{d\mathcal{E}_\lambda}{\lambda - \mu} \quad (\mu \in \text{reg } A).$$

It is easy to see from this that the spectrum of an operator $A$ coincides with the set of points of increase of the function\(^{52}\) $\mathcal{E}_\lambda$. The eigenvalues are the points of discontinuity of the function $\mathcal{E}_\lambda$. The operator $P_\lambda = \mathcal{E}_{\lambda+0} - \mathcal{E}_\lambda$ is an orthoprojection onto the corresponding eigenspace. The dimension of this subspace is called the *multiplicity* of the eigenvalue $\lambda$. The eigenvalues form the so-called *discrete spectrum* $\text{spec}_d A$. The *continuous spectrum* $\text{spec}_c A$ is defined as the complement in $\text{spec } A$ of the set of eigenvalues of finite multiplicity\(^{53}\). In terms of the resolution of the identity it can be characterised as the union of the set of non-isolated points of increase and the set of eigenvalues of infinite multiplicity. This characterisation combining these two sets is just the result that there exists a normalised quasi-eigensequence $(x_n)\(^{54}\) (\|x_n\| = 1, Ax_n - \lambda x_n \to 0)$ with no convergent subsequence. The intersection $\text{spec}_c A \cap \text{spec}_d A$ can turn out to be non-empty. However the whole space $\mathcal{E}$ can be decomposed into the orthogonal sum of the closed linear hull $E_d$ of the eigenvectors and the closed invariant subspace $E_c$ of vectors with purely continuous spectrum, i.e. those $x \in \mathcal{E}$ for which $(\mathcal{E}_\lambda x, x)$ is continuous. Then in turn $E = E_c \oplus E_{ac}$ where the summands are the spaces of *singular* and *absolutely continuous vectors* (these concepts are of course defined through the corresponding properties of the function $(\mathcal{E}_\lambda x, x)$). The extraction of the space $E_{ac}$ is particularly important for the theory of scattering (see the end of this section).

\(^{51}\) For rational $\phi$ with poles outside of $\text{spec } A$ this definition is equivalent to the previous one.

\(^{52}\) i.e. with the union of the sets of points of increase of all functions $(\mathcal{E}_\lambda x, x) (x \in \mathcal{E})$.

\(^{53}\) $\text{Spec}_c A$ and $\text{spec}_d A$ are defined for a normal (in particular a unitary) operator in exactly the same way.
$E_{ac} = E$ we say that $A$ is an operator with absolutely continuous spectrum and in the general case $A|_{E_{ac}}$ is called the absolutely continuous part of the operator $A$. Similar terminology can be introduced for the remaining types of spectra.

Of the other properties of the resolution of the identity corresponding to a selfadjoint operator $A$ we note that the measure $\int d\mathcal{E}_\lambda$ commutes with $A$, i.e. $\mathcal{E}_\Delta$ commutes with $A$ for any interval $\Delta = [\alpha, \beta]$. In fact

$$A\mathcal{E}_\Delta = \int_{\alpha}^{\beta} \lambda \, d\mathcal{E}_\lambda, \quad \mathcal{E}_\Delta A = A\mathcal{E}_\Delta|_{D(A)}.$$ 

One of the approaches to the spectral theorem for selfadjoint operators is through the Cayley transform

$$V = (A - iI)(A + iI)^{-1},$$ 

a unitary operator for which $\text{Im}(V - I)$ is dense. The point is that the spectral theorem also holds for unitary operators for which the proof can be effected by quite simple means.

**Theorem.** For any unitary operator $U$ there exists a (unique) orthogonal resolution of the identity $\mathcal{E}_\alpha$ ($0 \leq \alpha \leq 2\pi$) with

$$U = \int_{0}^{2\pi} e^{i\alpha} \, d\mathcal{E}_\alpha.$$ 

**Proof.** Let us consider the group of powers $(U^n)_{n=\infty}$. It is easy to verify that the Toeplitz matrix with elements $(U^{n-m}x, x)$ ($x \in E$) is positive. By the Riesz-Herglotz theorem there exists a non-decreasing function $\sigma(\alpha; x)$ ($0 \leq \alpha \leq 2\pi$) which is a solution of the trigonometric moment problem

$$(U^n x, x) = \int_{0}^{2\pi} e^{i\alpha n} \, d\sigma(\alpha; x) \quad (n = 0, \pm 1, \pm 2, \ldots). \quad (29)$$

Without loss of generality we can assume that $\sigma(\alpha; x)$ is continuous on the left on $(0, 2\pi)$ and that $\sigma(0; x) = 0$. Under these conditions $\sigma(\alpha; x)$ is unique and

$$\sigma(2\pi; x) = \int_{0}^{2\pi} d\sigma(\alpha; x) = (x, x).$$

Consequently $\sigma(\alpha; x) \leq \|x\|^2$ for all $\alpha$. Performing polarisation in (29), we obtain

$$\int_{0}^{2\pi} e^{i\alpha n} \, d\tau(\alpha; x, y) \quad (n = 0, \pm 1, \pm 2, \ldots), \quad (30)$$

where

$$\tau(\alpha; x, y) = \frac{1}{4} \sum_{k=0}^{3} \sigma(\alpha; x + i^k y) \quad (x, y \in E).$$

Since there is only one possible function $\tau$ in (30), it follows that $\tau(\alpha; x, y)$ is a sesquilinear functional of $x$, $y$ and hence $\tau(\alpha; x, y) = (\mathcal{E}_\alpha x, y)$, where $\mathcal{E}_\alpha$ is a
bounded selfadjoint operator. The required properties of the family \( \{ \mathcal{E}_a \} \) are verified without particular difficulty. \( \square \)

In order to pass to the selfadjoint operator \( A \) it is necessary to express it using the Cayley transform:

\[
A = i(V + 1)(V - 1)^{-1} = i \int_0^{2\pi} \frac{e^{i\alpha} + 1}{e^{i\alpha} - 1} d\mathcal{E}_a = \int_{-\infty}^{\infty} \lambda \, d\mathcal{E}_\lambda,
\]

where \( \mathcal{E}_\lambda = \mathcal{E}_{2\arccot \lambda} \).

There are several other proofs of the spectral theorem for a selfadjoint operator which are associated with different variants of the moment problem. For example, we can note that the resolvent \( R_\mu \) satisfies the inequality \( \text{im}(R_\mu x, x) > 0 \) \( (x \neq 0) \) in the half-plane \( \text{im} \mu > 0 \) and, as we already know,\(^{54}\)

\[
\|R_\mu\| \leq \frac{1}{|\text{im} \mu|} \quad (\text{im} \mu \neq 0).
\]

We can therefore make use of the integral representation of functions of the Nevanlinna class:

\[
(R_\mu x, x) = \int_{-\infty}^{\infty} \frac{d\omega(\lambda; x)}{\lambda - \mu},
\]

and it turns out as before that

\[
R_\mu = \int_{-\infty}^{\infty} \frac{d\mathcal{E}_\lambda}{\lambda - \mu},
\]

where \( \mathcal{E}_\lambda \) is an orthogonal resolution of the identity. Hence

\[
A = \mu 1 + R_\mu^{-1} = \int_{-\infty}^{\infty} \lambda \, d\mathcal{E}_\lambda.
\]

Now we have a further fruitful argument: the operator \( A \) is conservative, i.e. it generates a unitary group \( T(t) \) \( (t \in \mathbb{R}) \); since all the functions \( (T(t)x, x) \) \( (x \in E) \) are positive, \textit{Stone's theorem} that

\[
T(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \, d\mathcal{E}_\lambda
\]

follows from \textit{Bochner's theorem} and hence by differentiation at \( t = 0 \) we obtain a spectral resolution of the operator \( A \). Correspondingly \( T(t) = e^{i\lambda t} \).

If \( A \) is a densely defined closed symmetric operator with equal defect numbers, its resolutions of the identity are by definition generated by its selfadjoint extensions. For a non-selfadjoint operator \( A \) the resolution of the identity is not unique, therefore the question of its invariant (i.e. independent of the resolution) properties arises.

\(^{54}\) All this follows from the formula \( \text{im}(R_\mu x, x) = (\text{im} \mu) \|R_\mu x\|^2 \) which can be easily established with the help of Hilbert's identity.
Theorem. If the defect number \( n = n_+ = n_- \) of the operator \( A \) is finite, then all of its selfadjoint extensions have exactly the same continuous spectrum.

The general property of invariance of the continuous spectrum under 'small' perturbations which in some sense or other are finite rank or even compact is evident in this result and also in the following.

**Theorem of Weyl.** If \( A, B \) are selfadjoint operators with \( B \) compact, then \( \text{spec}_c(A + B) = \text{spec}_c(A) \).

In contrast to this the discrete spectrum is not invariant: for any selfadjoint operator \( A \) on a separable Hilbert space there exists a selfadjoint compact operator \( B \) with arbitrarily small norm such that the system of eigenvectors of the operator \( A + B \) is complete (von Neumann, 1935).

**Remark.** Let \( A, B \) be closed operators on a Banach space. The operator \( B \) is said to be \( A \)-compact if \( D(B) = D(A) \) and \( B|_{D(A)} \) is compact under the graph norm of the operator \( A \) (for this it is necessary and sufficient that the operator \( BR_A(A) \) be compact for some \( \lambda \) which is regular for \( A \)). For an \( A \)-compact operator \( B \) the set of points of the spectrum of the sum \( A + B \) that lie in any connected component \( G \) of the set \( \text{reg } A \) is at most countable (if it is countable then it condenses to \( \partial G \)). All these points are normal eigenvalues.

Let us consider certain classical situations from the point of view of the abstract spectral theorem.

**Example 1.** The operator \( \frac{1}{i} \frac{d}{ds} \) on \( L^2(-\pi, \pi) \) with periodic boundary conditions \( \phi(-\pi) = \phi(\pi) \) is selfadjoint. Its resolution of the identity is generated by the Fourier series expansion:

\[
(E_x \phi)(s) = \sum_{n < \lambda} c_n e^{ins}.
\]

The spectrum of the given operator consists of the eigenvalues \( \lambda_n = in \) and the corresponding eigenfunctions are \( e^{ins} \) \( (n \in \mathbb{Z}) \). This is an example of a purely discrete spectrum (there is no continuous spectrum).

**Example 2.** The resolution of the identity for the operator \( \frac{1}{i} \frac{d}{ds} \) on \( L^2(-\infty, \infty) \) has the form

\[
(E_x \phi)(s) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\lambda} \tilde{\phi}(\mu)e^{ins} d\mu,
\]

where \( \tilde{\phi} \) is the Fourier-Plancherel transform. The spectrum of the given selfadjoint operator fills out the whole axis, however there are no eigenvalues in it since

---

55 Hence it is clear that if \( B \) is compact then it is \( A \)-compact for any \( A \).
5.4. Theory of Operators

The functions $e^{i\mu x}$ do not belong to $L^2$. This is an example of a purely continuous spectrum.

**Example 3.** Let us consider the Schrödinger operator $-\frac{d^2}{d x^2} + q(x)y$ on $L^2(0, \infty)$ with continuous (for simplicity) real potential under the boundary condition $y'(0) - h y(0) = 0 (-\infty < h < \infty)$. First of all we define it only on the functions of finite support that are of class $C^2$ and satisfy the boundary condition; in this role we denote it by $Q_0$. The operator $Q_0$ is symmetric, densely defined but not closed. The domain of definition of the adjoint operator $Q_0^*$ consists of the functions $z \in L^2 \cap C^1$ such that $z'$ is absolutely continuous, $-z'' + q(x)z$, $z \in L^2$ and $z'(0) - h z(0) = 0$. Moreover $Q_0^* z = -z'' + q(x)z$. The domain of definition of the closure $\overline{Q}_0$ is extracted from $D(Q_0^*)$ by imposing the additional condition

$$\lim_{x \to \infty} [y'(x)z(x) - y(x)z'(x)] = 0 \quad (z \in D(Q_0^*))$$

This condition can be satisfied automatically (for example, this is the case if the potential is bounded from below) and then $\overline{Q}_0$ is a selfadjoint operator. In the contrary case the deficiency index of the operator $\overline{Q}_0$ is $(1, 1)$. This is precisely the Weyl alternative, since in the joint case the equation $Q_0^* z = \lambda z$ $(\im \lambda \neq 0)$ has only the trivial solution in $L^2$, while in the disk case it has a one-dimensional space of solutions. In any event a selfadjoint extension $Q$ of the operator $Q_0$ exists (it is unique in the first case but not in the second). All these extensions are in one to one correspondence with the spectral functions, namely, if $\rho(\lambda)$ is a spectral function then the resolution of the identity for the corresponding selfadjoint extension $Q$ is generated by the inversion formula associated with $\rho(\lambda)$:

$$(\mathcal{E}_\lambda f)(x) = \int_{-\infty}^{\lambda} F(\mu) \psi(x, \mu) \, d\rho(\mu)$$

(in the notation of Chapter 1, Section 2.12). From Parseval's identity

$$(\mathcal{E}_\lambda f, f) = \int_{-\infty}^{\lambda} |F(\mu)|^2 \, d\rho(\mu),$$

i.e. each measure $\int d(\mathcal{E}_\lambda f, f)$ is absolutely continuous with respect to the measure $\int d\rho(\lambda)$, its density being equal to the square of the modulus of the generalised Fourier transform of the function $f$. In this sense the resolution of the identity $\mathcal{E}_\lambda$ for the Schrödinger operator is absolutely continuous with respect to the measure $\int d\rho(\lambda)$. A similar result holds for the three-dimensional Schrödinger operator $-\Delta u + q(x)u$ on the whole space (A.Ya. Povzner, 1953) and for more general operators (Mautner, 1953; Yu.M. Berezanskij, 1956).

**Example 4.** Let us consider a Jacobian matrix $J$ with positive off-diagonal elements. Given an orthonormal basis $(e_i)_{i=0}^\infty$ of a Hilbert space $E$ the matrix $J$ defines a homomorphism $\tilde{J} : E \to s$ ($s$ is the space of all complex numerical sequences). Restricting $\tilde{J}$ to vectors of finite support we obtain a densely defined symmetric operator $J_0$ on $E$. It is not closed, therefore we have to pass to its
closure $\tilde{J}_0$. The adjoint operator $J_0^*$ is the restriction of the operator $\tilde{J}$ to vectors which satisfy the condition $\tilde{J}v \in E$. In accordance with the Weyl alternative the equation $J_0^*v = \lambda v$ either has only the trivial solution for all $\lambda$ (im $\lambda \neq 0$) and then the operator $\tilde{J}_0$ is selfadjoint (point case) or the space of solutions of this equation is one-dimensional for all $\lambda$ (im $\lambda \neq 0$) in which case the deficiency index of the operator $J_0$ is $(1, 1)$ (disk case). In any event a selfadjoint extension $A$ of the operator $J_0$ exists (it is unique in the first case but not in the second). All these extensions are in one to one correspondence with the extremal solutions $\sigma(\lambda)$ of the power moment problem associated with the given Jacobian matrix. In fact, if $\mathcal{E}_\lambda$ is the resolution of the identity for the operator $A$, then $\sigma(\lambda) = (\mathcal{E}_\lambda e_0, e_0)$.

Returning to the Schrödinger operator on the semiaxis, we note that progress in investigating its spectral properties was largely dependent on the discovery and development of the technique of so-called transformation operators (Delsarte, 1938; A.Ya. Povzner, 1948; B.M. Levitan, 1949) which establishes the equivalence of the operator $Q$ and the elementary operator $-y''$ with boundary condition $y'(0) = 0$. In this connection it is important that the transformation operator has the form $1 + K$, where $K$ is a Volterra integral operator with continuous kernel, so that the inverse operator has a similar form.

V.A. Marchenko (1950) applied transformation operators to a wide range of problems on the spectral analysis of the Schrödinger operator. In particular he obtained in this way the asymptotic expression for the spectral function

$$\rho(\lambda) \sim \frac{2}{\pi} \sqrt{\lambda} \quad (\lambda \to +\infty)$$

and established a theorem on the uniqueness of the solution to the inverse problem of reconstructing the operator from its spectral function. A linear procedure for solving this problem, which is also based on the use of transformation operators, was given by I.M. Gel'fand and B.M. Levitan (1951).

In the physically important inverse problem of scattering theory the input data (to begin with the so-called limiting phases) are linked with the asymptotic behaviour of the solutions of the equation $Qy = \lambda y$ as $x \to \infty$. An effective

---

57 In connection with generalised shift (g.s.). The simplest g.s. $(V(t)\phi)(s) = (\phi(s + t) - \phi(s - t))/2$ arises from the wave equation $\partial^2 u/\partial t^2 = \partial^2 u/\partial s^2$; the general theory of g.s. is in many respects parallel to the theory of group representations (B.M. Levitan, 1962).

58 It is a question of the construction of a certain linear integral equation for the kernel $K(x, \zeta)$, after which the formulae $q(x) = 2dK(x, x)/dx, h = K(0, 0)$ are applied. All this goes through under certain conditions on $\rho(\lambda)$ which are close to being necessary. Necessary and sufficient conditions for $\rho(\lambda)$ to be the spectral function of a Schrödinger operator on the semiaxis with continuous potential were obtained in a different way by M.G. Krejn (1951).

59 Recently the inverse problem of scattering theory has been at the centre of attention of many investigations thanks to its utilisation for integrating the Korteweg-de Vries equation and other non-linear equations of mathematical physics.

60 The description of which forms the direct problem. We note that both problems (direct and inverse) under consideration are concerned with the stationary theory of scattering, which is formulated in terms associated with the energy spectrum and not with evolution in time.
method for its solution (V.A. Marchenko, 1955) therefore required the use of transformation operators attached to infinity:

$$\phi \mapsto \phi(x) + \int_x^\infty K(x, \xi)\phi(\xi)\,d\xi.$$}

The existence of such operators for potentials which decrease sufficiently rapidly as $x \to \infty$ was established by B.Ya. Levin (1955).

The inverse problem of scattering theory on the whole axis was solved by L.D. Faddeev (1964), also on the basis of the technique of transformation operators.

**Remark.** For differential operators of order $n > 2$ transformation operators of the form $1 + K$, where $K$ is a Volterra integral operator, do not in general exist (V.I. Matsaev, 1960), however in the case of analytic coefficients the transformation operator was constructed by L.A. Sakhnovich (1961).

In **non-stationary scattering theory** the problem is to compare the asymptotic behaviour as $t \to +\infty$ of a physical system which is evolving in time $t$ with its behaviour as $t \to -\infty$.

**Example.** The Schrödinger operator $Q$ (just as any other selfadjoint operator) generates a group of unitary operators $e^{itQ}$. In quantum mechanics there corresponds to this a **non-stationary Schrödinger equation** for the wave function $\psi$:

$$ih\frac{\partial \psi}{\partial t} = H\psi,$$

where $H$ is a Hamiltonian (energy operator), $\hbar = \frac{h}{2\pi}$ and $h$ is Planck's constant.

If a particle with mass $m$ is in a force field with potential $q(x)$ ($x \in \mathbb{R}^3$), then

$$H\psi = -\frac{\hbar^2}{2m}\Delta \psi + q(x)\psi$$

(for example, this represents an $\alpha$-particle in the field of a massive nucleus which was the situation in the historical experiments of Rutherford). The effect of scattering is created by the potential $q$, i.e. it is associated with the deviation of the actual evolution from the free (unperturbed) evolution which is determined by the Hamiltonian

$$H_0\psi = -\frac{\hbar^2}{2m}\Delta \psi.$$}

This deviation is described by the **wave operators**

$$W_\pm = \lim_{t \to \infty} e^{-iHt}e^{iH_0t},$$

which are clearly isometric. The limits here are naturally treated as strong limits.

---

\[62\] This is two-particle scattering. $N$-particle scattering with $N > 2$ is a much more complicated process. It was investigated for $N = 3$ by L.D. Faddeev (1963).
and the first question concerns their existence (if not on the whole of \(L^2\) then on subspaces which have to be described). Following that we can define the \textit{scattering operator} \(S = (W_+ W_-)^{-1} W_-\),

but this construction is satisfactory only if \(\text{Im } W_- = \text{Im } W_+\) since then \(S\) maps \(D(W_-)\) isometrically onto \(D(W_+)\). It is even better if we also have \(D(W_-) = D(W_+)\), since then \(S\) is a unitary operator on a certain Hilbert space. The elucidation of these cases is the second important question of scattering theory.

Clearly the stated questions can be posed in a general form, i.e. for a pair of selfadjoint operators \(A_0, A\) on a Hilbert space. Such an 'abstract' theory is capable of enveloping both the quantum and the classical scattering processes (acoustic, optical etc.). The non-trivial nature of the problem is underlined by the fact that, if \(A\) is a selfadjoint operator for which zero is not an eigenvalue, then \(e^{iAt}\) \((t \neq 0)\) cannot have a limit under the norm as \(t \to +\infty\) (as also for \(t \to -\infty\)). This also prevents the existence of wave operators in the presence of a discrete spectrum for \(A_0\). The rigorous theory of scattering is concentrated on the absolutely continuous spectrum.

\textbf{Kato-Rosenblum Theorem.} If \(A = A_0 + B\) where \(B\) is a nuclear operator, the wave operators \(W_\pm\) are defined on the subspace \(E_{ac}(A_0)\) and map it onto \(E_{ac}(A)\). Thus the scattering operator \(S\) is defined and is unitary on \(E_{ac}(A_0)\).

Each of the operators \(W_\pm\) intertwines the representations

\[ e^{iA_0t}|_{E_{ac}(A_0)}, \quad e^{iAt}|_{E_{ac}(A)}. \]

Therefore the absolutely continuous parts of the operators \(A_0, A\) are unitarily equivalent.

In applications the conditions of the Kato-Rosenblum theorem (1957) are not usually satisfied, however Kuroda (1960) discovered more effective conditions for this purpose. They in turn are covered by a \textit{theorem of M.S. Birman and M.G. Krejn} (1962), in which the resolvents of the operators \(A_0, A\) are required to be nuclear. It is even sufficient that the difference of the resolvents of \(\phi(A_0)\) and \(\phi(A)\) be nuclear, where the function \(\phi\) belongs to a certain sufficiently wide class (M.S. Birman, 1962).

\textbf{4.9. Spectral Operators.} Functions of a selfadjoint operator are normal operators. However the spectrum of a normal, for example unitary, operator is not in general real. If

\[ \Phi = \int_{-\infty}^{\infty} \phi(\lambda) \, d\mathcal{E}_\lambda, \]

\(63\) The \textit{Heisenberg S-matrix} is unitarily equivalent to this operator by means of the Fourier transform \(\psi \mapsto \hat{\psi}\) (i.e., from the physical point of view, the transition from the coordinate space into the impulse space).

\(64\) The case of commuting \(A, A_0\) is not of interest in scattering theory.
where $\mathcal{E}_\mu$ is an orthogonal resolution of the identity, then $\text{spec } \Phi$ is the closure of the set of essential values\(^{65}\) of the function $\phi$ (spectral mapping theorem). The spectral resolution (31) is written more naturally in the form

$$\Phi = \int_{\mathbb{C}} \mu \, d\mathcal{E}_\mu,$$

where $\int d\mathcal{E}_\mu$ is the operator-valued measure in the complex plane $\mathbb{C}$ which is the image of the measure $\int d\mathcal{E}_\lambda$ under the mapping $\phi$. The support of the measure in (32) coincides with the spectrum of the operator $\Phi$.

The spectral resolution of a normal operator in the form (32) provided the starting point of the theory of spectral operators on a Banach space which was established by Dunford and Bade in the 1950s. In this theory it is postulated that there exists a resolution of the identity $\mathcal{E}(M)$ which is a projection-valued measure on $\mathbb{C}$ ($\mathcal{E}(1) = 1$), commutes with a given closed densely defined operator $A$ and is such that $\text{spec}(A|_{\mathbb{C} \mathbb{M}(M)}) = M$ (in this case the operator $A$ is said to be spectral). The resolution of the identity for any spectral operator is unique and if $A$ is bounded\(^{66}\) then

$$A = \int_{\mathbb{C}} \lambda \mathcal{E}(d\lambda) + R,$$

where $R$ is a quasinilpotent operator which commutes with $A$. The term

$$A = \int_{\mathbb{C}} \lambda \mathcal{E}(d\lambda)$$

is also a spectral operator with the same resolution of the identity, $\text{spec } A = \text{spec } A$ and moreover the whole spectrum is approximate.

We say that operators of the form (33) are of spectral type. Therefore $A$ is called the scalar part of the operator $A$ and $R$ is its quasinilpotent part. The representation of a bounded spectral operator as the sum of an operator of scalar type and a quasinilpotent operator is unique.

The class of spectral operators is rather narrow. For example, the shift operator on the two-sided $l^p$ space ($1 \leq p \leq \infty; p \neq 2$) is not spectral (Fixman, 1959). The operator $\frac{1}{i} \frac{d}{ds}$ on $L^p(-\infty, \infty)$ or on $L^p(-\pi, \pi)$ with periodic boundary conditions also fails to be spectral if $p \neq 2$. We shall describe below a much wider class of operators (those with so-called separated spectrum).

4.10. Spectral Subspaces. Let $A$ be a closed densely defined operator on a Banach space $E$. A closed subspace $L \subset E$ ($L \neq 0$) is said to be spectral for $A$ if

---

\(^{65}\) A value $\lambda$ is inessential if $\phi(\lambda) \neq \lambda$ almost everywhere with respect to each measure $m_\mu = \int d\mathcal{E}_\mu(x, x)$.

\(^{66}\) For unbounded spectral operators the situation is considerably more complicated than that described below.
1) $L$ is invariant and $D(A) \cap L$ is dense in $L$, 2) $\text{spec}(A|_L) \subset \text{spec} A$, 3) if a closed invariant subspace $M \neq 0$ is such that $D(A) \cap M$ is dense in $M$ and $\text{spec}(A|_M) \subset \text{spec}(A|_L)$ then $M \subset L$. We shall call a compact set $Q \subset \text{spec} A$ spectral if there exists a spectral subspace $L = L(Q)$ for which $\text{spec}(A|_L) = Q$ (this subspace is clearly unique). As a trivial example we have: if $A$ is bounded then the compact set $\text{spec} A$ is spectral, the corresponding spectral subspace $L$ being $E$. It can happen that no other spectral compact sets exist.

**Example.** Let us consider the Banach space of scalar functions which are analytic on the disk $|\lambda| < 1$ and continuous up to its boundary under the usual norm $\max_{|\lambda|=1} |\phi(\lambda)|$. We have on it the operator $A$ of multiplication by $\lambda$. The spectrum of the operator $A$ coincides with the disk $D = \{ \lambda : |\lambda| \leq 1 \}$. Suppose that $Q$ is a non-trivial spectral compact set with corresponding spectral subspace $L$. We put $A = D\setminus Q$. If $\mu \in \Delta$, for any $\phi \in L$ the function $(\lambda - \mu)^{-1} \phi(\lambda)$ is regular for $\lambda = \mu$, but then $\phi(\mu) = 0$. Thus if $\phi \in L$ we have $\phi|_\Delta = 0$. It then follows from the uniqueness theorem that $\phi = 0$, i.e. $L = 0$ in spite of the definition of a spectral subspace.

Any compact set $Q \subset \text{spec} A$ which is open in $\text{spec} A$ is spectral. The corresponding spectral subspace is obtained as the image of the projection

$$P = -\frac{1}{2\pi i} \int_\Gamma R_\lambda d\lambda,$$

where $\Gamma$ is a simple closed contour separating $Q$ from its complement $Q'$ in $\text{spec} A$. Moreover $1 - P$ is the projection onto the spectral subspace corresponding to the compact set $Q$. This construction is contained in the *functional calculus of F. Riesz*, which we shall describe below under the assumption that $A$ is bounded. Let us put

$$\phi(A) = -\frac{1}{2\pi i} \int_\Gamma \phi(\lambda) R_\lambda d\lambda,$$  \hspace{1cm} (34)

where $\phi$ is a function which is analytic on $\text{spec} A$, i.e. on some neighbourhood $G \supset \text{spec} A$, and $\Gamma$ is a closed contour which does not cross itself, lies in $G$ and contains $\text{spec} A$ in its interior. Formula (34) defines a homomorphism of the algebra of functions which are analytic on $\text{spec} A$ into the algebra of bounded operators, such that if $\phi(\lambda) = \lambda$ then $\phi(A) = A$.

---

67 This term requires a more precise definition since $A$ need not be defined everywhere on $E$. Invariance of a subspace $L$ means that if $x \in D(A) \cap L$ then $Ax \in L$. The restriction $A|_L$ has to be interpreted as the restriction of $A$ to $D(A) \cap L$. It is clearly closed along with $A$, however $D(A) \cap L$ may fail to be dense in $L$ when $D(A)$ is dense.

68 Thus the spectral subspace $L = \text{Im} P$ is completely reducing. If $Q$ consists of a single point $\mu$ and $L$ is finite-dimensional, then $\mu$ is an ordinary eigenvalue and $L$ is the corresponding maximal root subspace.

69 The neighbourhood mentioned and the contour $\Gamma$ can be disconnected. The integral does not depend on the choice of contour.
For the calculus just described, Dunford (1943) established the spectral mapping theorem, spec $\phi(A) = \phi(\text{spec } A)$, and the superposition theorem: $(\psi \circ \phi)(A) = \psi(\phi(A))$ under the condition that $\phi$ is analytic on spec $A$ and $\psi$ is analytic on $\phi(\text{spec } A)$.

If the function $\phi$ is analytic on the disk $|\lambda| < r$, where $r > \rho(A)$, then, expanding it in a Taylor series and using the expansion (23), we obtain from (34)

$$\phi(A) = \sum_{n=0}^{\infty} \frac{\phi^{(n)}(0)}{n!} A^n,$$

i.e. we arrive at the previous definition of the function from the operator.

We shall say about an operator $A$ on a Banach space $E$ that its spectrum is \textit{separated} if each compact set $Q \subset \text{spec } A$ which is the closure of its set of interior points relative to spec $A$ is spectral. For example, any spectral operator is of this type: the spectral subspace $L(Q)$ is the image of the projection $\pi_Q = s d\lambda$.

Separatedness of the spectrum of an operator is ensured by a certain bound on the growth of the resolvent close to the spectrum. In fact, suppose that spec $A$ lies on a curve $C \subset \mathbb{C}$ and that in a neighbourhood of each point $\mu \in \text{spec } A$ the inequality

$$\|R_{\lambda}\| \leq M_{\mu}(d(\lambda, C)) \quad (\lambda \notin C),$$

is satisfied, where $M_{\mu}(\delta)$ is a decreasing function of $\delta > 0$ ($\delta \leq d_0 = d_0(\mu)$) such that the Levinson condition

$$\int_{0}^{d_0} \ln \ln M_{\mu}(\delta) \, d\delta < \infty$$

is satisfied. Then $A$ is an operator with separated spectrum (Yu.I. Lyubich – V.I. Matsaev, 1960). In the proof of this theorem a special functional calculus is used, with the help of which \textit{quasiprojections71} onto the spectral subspaces are constructed. It also allows us to establish (under conditions (35), (36)) the completeness of the system of spectral subspaces72 corresponding to any covering of the spectrum by finite arcs of the curve $C$.

\textbf{Example.} The Schrödinger operator on $L^2(-\infty, \infty)$ with periodic complex potential has separated spectrum (V.A. Tkachenko, 1964).

For a bounded operator $A$ with real spectrum Levinson's condition is equivalent to Ostrowski's condition with respect to the weight $\alpha(t) = \|e^{itA}\| (t \in \mathbb{R})$:

---

70 This can be developed further in the case of a spectral operator.
71 These are operators $\Phi$ associated with a compact set $Q$ and a neighbourhood $U \ni Q$ such that $\Phi|_{L_{UQ}} = 1$ and $\Phi|_{L_{UK}} = 0$ for compact sets $K$ lying outside of $U$.
72 A certain theory of duality also holds for spectral subspaces under conditions of type (36) (Bishop, 1959; Yu. I. Lyubich, V.I. Lomonosov, V.I. Matsaev, 1973).
The latter signifies the non-quasianalyticity\(^73\) of the Fourier image of the class \(L_1^1(\mathbb{R})\) of functions which are summable with respect to the weight \(a\). Starting from (37\(_R\)) it is possible to construct the functional calculus

\[
\tilde{\phi}(A) = \int_{-\infty}^{\infty} \phi(t) e^{-iAt} \, dt \quad (\phi \in L_1^1(\mathbb{R}))
\]

independently of the previous approach; this is the operator Fourier transform, which allows us to obtain functions of compact support of the operator \(A\), in particular quasiprojections onto spectral subspaces. Moreover \(L(Q)\) is constructed as the intersection of all the subspaces

\[\{x: \tilde{\phi}(A)x = x\} \quad (\tilde{\phi}|_Q = 1).\]

The requirement of boundedness of the operator \(A\) is not essential, the important thing being rather that it generates a group\(^74\) \(T(t) (t \in \mathbb{R})\). The stated conditions of growth of the resolvent or exponential cannot be weakened: under any more rapid growth rate there arise operators for which the spectrum on any closed invariant subspace coincides with the spectrum on the whole space.

For a bounded operator \(V\) with spectrum on the unit circle the requirement

\[
\sum_{n=-\infty}^{\infty} \ln \frac{\|V^n\|}{n^2 + 1} < \infty \quad (37\_T)
\]

is the analogue of Ostrowski's condition and it is equivalent to Levinson's condition; thus separatedness of the spectrum of the operator\(^75\) \(V\) follows from this. In particular, any isometric operator on a Banach space has separated spectrum but it is certainly not the case that any such operator is spectral. Any closed subset \(F\) of the circle (or line in the case of (37\(_R\))) is the spectrum of some non-quasianalytic operator\(^76\). If the system of eigenvectors of a non-quasianalytic operator is complete, its spectrum \(S\) coincides with the closure of its set \(\mathbb{S}\) of eigenvalues (N.K. Nikolskij, 1971), since in the contrary case there exists a compact set \(\mathcal{Q} \subset S\) such that \(\mathcal{Q} \cap \overline{\mathbb{S}} = \emptyset\). But then there is a quasiprojection which annihilates all the eigenvectors and is the identity on \(L(\mathcal{Q})\).

A bounded operator \(A\) on a Banach space \(E\) is said to be **decomposable** if for any open covering \(\bigcup_{k=1}^{n} G_k \supset \text{spec} \, A\) there exists a system of spectral subspaces

\(73\) An operator \(A\) which satisfies condition \((37\_R)\) is said to be non-quasianalytic.

\(74\) In this general case the condition \((37\_R)\) is stronger than (36).

\(75\) Under condition \((37\_T)\) the operator, as before, is said to be non-quasianalytic.

\(76\) Any non-empty closed subset \(F \subset \mathbb{C}\) is the spectrum of some normal operator (selfadjoint if \(F \subset \mathbb{R}\), unitary if \(F \subset \mathbb{T}\) (the unit circle)), namely the operator of multiplication by the independent variable on the space \(L^2(F, \mu)\), where \(\mu\) is any measure with support \(F\). As regards this we note that a bounded set \(A \subset \mathbb{C}\) is the set of all eigenvalues of a bounded operator on a separable Hilbert space if and only if it is of type \(F_\mu\) (L.N. Nikol'skaya, 1970).
such that \( \text{spec}(A|_{L_k}) \subseteq G_k \) \((1 \leq k \leq n)\) and \( \sum_{k=1}^{n} L_k = E \). The theory of this class of operators has been developed in depth during the last 25 years in the works of Foias and his school.

4.11. Eigenvectors of Conservative and Dissipative Operators. Completeness of the system of eigenvectors (eigenspaces) of a conservative operator \( A \) on any Banach space \( E \) is equivalent to the almost-periodicity of the corresponding isometric representation

\[ T(t) \quad (t \in \mathbb{R}). \]

The necessity of almost-periodicity is obvious: if \( \{x_k\}_{1}^{n} \) is a collection of eigenvectors corresponding to the eigenvalues \( \{i\lambda_k\}_{1}^{n} \) and \( \|x - \sum_{n} x_n\| < \varepsilon \), then

\[ \|T(t)x - \sum_{k=1}^{n} x_k e^{i\lambda_k t}\| < \varepsilon \quad (t \in \mathbb{R}). \]

Hence it is clear that the orbit of the vector \( x \) is relatively compact.

The sufficiency of almost-periodicity of the representation \( T \) for completeness of the system of eigenvectors of the operator \( A \) can be established with the help of the theory of representations of compact groups. In particular, the spectral theorem for compact selfadjoint operators is generalised by this method: if \( A \) is a compact conservative operator on a reflexive Banach space \( E \) then its system of eigenvectors is complete (Yu.I. Lyubich, 1960). The reflexivity condition can be relaxed: it is sufficient that \( E \) should contain no subspace isomorphic to \( c \). Moreover there is a counterexample in \( c \) (Yu.I. Lyubich, 1963).

Example. Let \((\lambda_n)_{0}^{\infty}\) be a sequence of distinct positive numbers which converges to zero. Let us consider the diagonal operator

\[ A(\xi_n) = (i\lambda_n \xi_n) \]

on \( c \). It is compact and conservative:

\[ e^{2\pi it}A(\xi_n) = (e^{2\pi it} \xi_n). \]

However its eigenvectors \( \delta_m = (\delta_m)_{n=0}^{\infty} \) \((m = 0, 1, 2, \ldots)\) only generate the subspace \( c_0 \) of sequences which converge to zero.

Remark. The ergodic theorem for an almost-periodic representation \( T(t) \) \((t \in \mathbb{R})\) follows easily from the completeness of the system of eigenvectors of the generator \( A \): the existence of the limit

\[ \lim_{s \to \infty} \frac{1}{2s} \int_{-s}^{s} (T(t)x) \, dt \]

for each eigenvector \( x \) is clear, i.e. \( T(t)x = e^{iut}x \) for \( Ax = i\mu x \).

---

77 By the same token this is also valid for any bounded representations:

\[ \sup_{t} \|T(t)\| = M < \infty. \]

78 We note that this approach also leads to the classical theorem of Bohr for almost-periodic functions – the case of a regular representation in \( AP(\mathbb{R}) \).
According to the ergodic theorem the strong limit

\[ P_\lambda = \lim_{s \to \infty} \int_{-s}^{s} T(t) e^{-i\lambda t} \, dt \]

exists for each \( \lambda \in \mathbb{R} \). This is a projection (\( \| P_\lambda \| \leq M \), therefore if \( T \) is isometric, \( P_\lambda \) is an orthoprojection) on \( W_\lambda = \ker(A - i\lambda I) \) and also \( P_\lambda P_\mu = 0 \), i.e. \( P_\lambda|_{W_\mu} = 0 \) (\( \lambda \neq \mu \)). Corresponding to each vector \( x \in E \) we have a Fourier series in terms of eigenvectors of the operator \( A \):

\[ x \sim \sum_{\lambda} P_\lambda x \]

(in the case of a regular representation in \( AP(\mathbb{R}) \) this is the classical Fourier series of an a.p.f.) The uniqueness theorem holds: \( \forall \lambda: P_\lambda x = 0 \Rightarrow x = 0 \). The proof reduces to the scalar case if we consider the a.p.f. \( f(T(t)x) (f \in E^*) \) since the Fourier coefficients of this function are \( c_\lambda = f(P_\lambda x) \).

Using the Fejér-Bochner method of summation for Fourier series of a.p.f.s we can associate with each subgroup \( \Gamma \subseteq T \) a projection \( P_\Gamma \) (with bound \( \| P_\Gamma \| \leq M \)) on the closure of the sum of the eigenspaces corresponding to eigenvalues \( \lambda \in \Gamma \) (N.K. Nikolskij, 1971). The replacement of subgroups by arbitrary subsets is not possible (Rudin, 1962).

Thus once again we have evidence of a close connection between harmonic analysis and the spectral theory of operators. There is however another approach to the problem of completeness of the system of eigenvectors of a conservative compact operator, which is based on an analytic technique using a bound for the resolvent (even weaker than (26)).

We say that an operator \( A \) on a Banach space \( E \) is quasiconservative if \( \text{spec } A \) is purely imaginary and

\[ \| R_\lambda \| \leq \frac{M}{|\text{re } \lambda|} \quad (\text{re } \lambda \neq 0, M \text{ a positive constant}). \]

If this inequality is satisfied for \( \text{re } \lambda > 0 \) (spec \( A \) lies in the half-plane \( \text{re } \lambda \leq 0 \)) we say that the operator is quasidissipative.

It is clear that in a Hilbert space each operator of the form \( iA \), where \( A \) is similar to a selfadjoint operator, is quasiconservative but the class of quasiconservative operators is not exhausted by these (A.S. Markus, 1966). Moreover even in a Banach space this class has an interesting spectral theory. Its simplest manifestation can be seen in the result that if an operator is conservative then the order of each eigenvalue is equal to one, i.e. there are no root vectors other than eigenvectors. It can be shown further that a quasinilpotent quasiconservative operator must be zero, because in a reflexive Banach space \( E \) the system of eigenvectors of a compact quasiconservative operator turns out to be complete and moreover (even without the requirement of compactness) there exists a system of uniformly bounded projections \( P_\lambda(\| P_\lambda \| \leq M, P_\lambda P_\mu = 0 \text{ if } \lambda \neq \mu) \) onto

\[ \{ i\lambda: P_\lambda \neq 0 \} \text{ coincides with the set of eigenvalues.} \]
the eigenspaces. Hence it follows that any normalised system \( \{e_\lambda\} \) of eigenvectors corresponding to pairwise distinct eigenvalues is distal:

\[
\|e_\lambda - e_\mu\| \geq \frac{1}{M} \|P_\lambda(e_\lambda - e_\mu)\| = \frac{1}{M}.
\]

Consequently the power of the set of eigenvalues is no more than \( \dim E \).

Generally speaking for operators related to selfadjoint operators we can expect a properly organised' system of eigenvectors to have properties close to those of an orthonormal system.

**Mukminov-Glazman Theorem.** Let \( iA \) be a bounded dissipative operator on a Hilbert space and \( \{\lambda_n\}_1^\infty \) a subset of its set of eigenvalues such that

\[
\sum_{m \neq n} \frac{(\text{im } \lambda_m)(\text{im } \lambda_n)}{|\lambda_m - \lambda_n|^2} < \infty. \tag{38}
\]

Then a system \( \{u_n\}_1^\infty \) of corresponding normalised eigenvectors is a Bari basis of its closed linear hull \( E_0 \).

**Proof.** Without loss of generality we can assume that the sum of the series (38) is equal to \( q^2 < \frac{1}{4} \). Since \( \text{im}(A(au_m + \beta u_n), au_m + \beta u_n) \geq 0 \) for all \( \alpha, \beta \), we have

\[
|u_m, u_n|^2 \leq \frac{4(\text{im } \lambda_m)(\text{im } \lambda_n)}{|\lambda_m - \lambda_n|^2} \quad (m \neq n).
\]

Hence by (38)

\[
\sum_{m, n=1}^\infty |u_m, u_n| - \delta_{mn}|^2 < \infty.
\]

In addition

\[
|\sum_{n=1}^N c_n u_n|^2 \geq (1 - q) \sum_{n=1}^\infty |c_n|^2,
\]

from which it follows that the system \( \{u_n\}_1^\infty \) is \( \infty \)-linearly independent. Thus this system is a Bari basis in \( E_0 \) by Krein's criterion. \( \square \)

The connection between selfadjoint and unitary operators on a Hilbert space, which is realised by means of the Cayley transform, extends to dissipative operators and contraction operators correspondingly. If \( iA \) is a dissipative operator its Cayley transform

\[
V_\lambda = (A - \lambda \mathbf{1})(A - \overline{\lambda} \mathbf{1})^{-1} \quad (\text{im } \lambda > 0)
\]

is a contraction and also \( \text{Im}(V_\lambda - \mathbf{1}) = D(A) \) is dense. Conversely, any contraction \( V \) for which \( \text{Im}(V - \mathbf{1}) \) is dense (since \( \text{Ker}(V - \mathbf{1}) = 0 \)) is the Cayley transform of the dissipative operator \( iA = (\lambda V - \lambda \mathbf{1})(V - \mathbf{1})^{-1} \). This remark allows us to transform theorems on dissipative operators into analogous theorems on contractions and conversely. Thus, for example, we have the analogue of the Mukminov-Glazman theorem for contractions with condition
\[
\sum_{m \neq n} \frac{(1 - |\lambda_m|^2)(1 - |\lambda_n|^2)}{|1 - \lambda_m \lambda_n|^2} < \infty, \tag{39}
\]

which results from replacing \(\lambda_m, \lambda_n\) in (38) by their images with respect to the fractional linear transformation \(\lambda = -i(\zeta + 1)(\zeta - 1)^{-1}\).

Remark. Condition (39) is only meaningful for eigenvalues which lie in the disk \(|\lambda| < 1\). This is not surprising if we take account of the theorem of Szőkefalvi-Nagy, Foias and Langer on the decomposition of an arbitrary contraction into an orthogonal sum of a unitary operator and a completely non-unitary (i.e. it has no unitary parts) operator.

A similar remark can be made with regard to dissipative operators.

Conditions (38) and (39) can be expressed in terms of Blaschke products:

\[
(38) \Leftrightarrow \prod_{m \neq n} \frac{|\lambda_m - \lambda_n|}{|\lambda_m - \lambda_n|} > 0; \quad (39) \Leftrightarrow \prod_{m \neq n} \frac{|\lambda_m - \lambda_n|}{1 - \lambda_m \lambda_n} > 0.
\]

The weaker conditions

\[
\inf_{m} \prod_{m \neq n} \frac{|\lambda_m - \lambda_n|}{|\lambda_m - \lambda_n|} > 0; \quad \inf_{m} \prod_{m \neq n} \frac{|\lambda_m - \lambda_n|}{1 - \lambda_m \lambda_n} > 0 \tag{40}
\]

arise in the problem investigated by Carleson (1958) of the interpolation \(\varphi(\lambda_n) = \alpha_n (n = 1, 2, 3, \ldots)\) of bounded sequences \((\alpha_n)\infty_{n=1}\) by functions \(\varphi\) of class \(H^\infty\) which are analytic and bounded on the half-plane in \(\lambda > 0\) or, respectively, on the disk \(|\lambda| < 1\). In fact these conditions are necessary and sufficient for solvability of the stated problem for all \((\alpha_n)\infty_{n=1} \in l^\infty\).

Starting from Carleson's criterion, V. Eh. Katsnel'son proved that if in the Mukminov-Glazman theorem condition (38) is weakened to the corresponding condition in (40), then the system \(\{u_n\}\infty_{n=1}\) will be a Riesz basis of its closed linear hull (and the situation is similar for contractions).

The proof (for a contraction \(V\)) is based on the fact that for any finite set \(S\) of natural numbers a function \(\varphi \in H^\infty\) can be constructed on the disk \(|\lambda| < 1\) such that \(\varphi(\lambda_n) = 1 (n \in S), \varphi(\lambda_n) = 0 (n \notin S)\) and \(\|\varphi\| = \sup |\varphi| \leq C\), where \(C\) does not depend on the choice of \(S\). If only a finite set of interpolation conditions is retained \((n \leq N)\), the function \(\varphi\) can be replaced by a rational \(\varphi_N\) (in accordance with a theorem of Schur), after which a theorem of von Neumann can be applied to show that \(\|\varphi_N(V)\| \leq \|\varphi_N\|\). As \(N \to \infty\) the sequence of operators \((\varphi_N(V))\) converges strongly to a projection \(P_S\) onto \(\text{Lin}\{u_n\}_{n \in S}\) along \(\text{Lin}\{u_n\}_{n \notin S}\). Since \(\|P_S\| \leq C\) for all \(S\), we have that \(\{u_n\}\infty_{n=1}\) is an unconditional basis in its closed linear hull.

---

\(^{80}\) With subsequent return to the previous notation.

\(^{81}\) Correspondingly (38) applies to eigenvalues which are not real.

\(^{82}\) For an isometric operator \(V\) this is the so-called Wold decomposition; moreover in this case the completely non-unitary part turns out to be a one-sided shift operator, i.e., it acts on a space of type \(L \oplus VL \oplus V^2L \oplus \cdots\), where \(L\) is some suitable subspace (and the unitary part acts on \(\bigcap_{n=1}^\infty V^nE\)).
4.12. Spectral Sets and Numerical Ranges. The theorem of von Neumann which has been mentioned is concerned with an aspect of spectral theory which was discovered by him; this is connected with subtle variants of functional calculus. We will discuss this topic briefly. Let $A$ be an operator on a Banach space and $S$ a closed subset of the complex plane $C$ which contains $\text{spec } A$. The set $S$ is said to be spectral for $A$ if the bound $\|\varphi(A)\| \leq \sup_S |\varphi|$ holds for all rational functions $\varphi$ whose poles lie outside of $S$. Von Neumann's theorem asserts that the disk $|\lambda| \leq 1$ is a spectral set for all contractions on a Hilbert space.\(^8\)

This is false on a Banach space,\(^9\) however, as V. Eh. Katnelson and V.I. Matsaev (1967) showed, it becomes true for the disk $|\lambda| \leq 3$ and the radius $\rho = 3$ cannot be decreased for the class of all Banach spaces.

For a normal operator on a Hilbert space its own spectrum is a spectral set (in consequence of the spectral theorem). Conversely, if any function which is continuous on $(\text{spec } A) \cup \{\infty\}$ can be approximated uniformly on this set by rational functions, then the operator $A$ is normal.

Another useful substitute for the spectrum of an operator $A$ is its so-called numerical range. In the case of a Hilbert space it is defined as the closure of the set of values of the quadratic functional $(Ax, x)$ for $x \in D(A), \|x\| = 1$. According to a theorem of Hausdorff the numerical range of any operator $A$ is convex.\(^5\) It contains the approximate spectrum and consequently the convex envelope of this set, but in general the numerical range is more extensive than the convex envelope of $\text{spec } A$. For a normal operator these two sets coincide.

The numerical range of an operator $A$ on a Banach space is defined to be the closure of the set of values $f_x(Ax)$, where $x \in D(A), \|x\| = 1$ and $f_x$ is a support functional at the point $x$ (the mapping $x \mapsto f_x$ is said to be supporting, as in the smooth case, but now, in general, it is not uniquely defined). The following example (Bonsall-Duncan, 1980) shows that the numerical range is not generally convex: for the space $C(S)$ ($S$ compact) with $f_x = \delta_{s(x)}$, where $s(x)$ is any point at which $\max_S |x(s)|$ is attained, the numerical range of the operator defined by multiplication by a function $\lambda(s)$ coincides with the set of values of this function.

The numerical range depends on the choice of support mapping but, if $A$ is a bounded operator ($D(A) = E$), all its numerical ranges have the same convex envelope (Lumer-Phillips, 1961). This set coincides with the intersection of the half-planes Re($\zeta e^{-i\theta}$) $\leq \tau(\theta)$, where $\tau(\theta)$ ($0 \leq \theta < 2\pi$) is the minimum of those $\tau$ for which $\|e^{ix\zeta}\| \leq e^{x\tau}$ (arg $\zeta = \theta$), i.e. $\|R_\zeta\| \leq 1/(|\zeta| - \tau(\theta))$ (arg $\zeta = \theta$, $|\zeta| > \tau(\theta)$). The convex envelope of the spectrum is contained in the numerical range.

In all cases the numerical range of an operator $A$ contains its approximate spectrum.

---

\(^8\) Conversely, if the unit disk is a spectral set for an operator $A$ (even on a Banach space) then $A$ must be a contraction. This follows trivially from the definition with $\varphi(\lambda) = \lambda$.

\(^9\) If the unit disk is a spectral set for all contractions on some Banach space then the space must be Hilbert (Foias, 1957).

\(^5\) It is enough to establish this result in 2-dimensional Euclidean space. But in this case the numerical range is an ellipse (which can degenerate into an interval) or a point (for a scalar operator).
For any choice of support mapping the numerical range of a dissipative operator lies in the half-plane $\Re \lambda \leq 0$. Conversely, if the numerical range of a closed, densely defined operator lies in the half-plane $\Re \lambda \leq 0$, the half-plane $\Re \lambda > 0$ is quasiregular and if in addition $n_+ = 0$ the operator is dissipative.

4.13. Complete Compact Operators. A compact operator $A$ on an LTS $E$ is said to be complete if the system $\Sigma_0(A)$ of root vectors corresponding to the non-zero eigenvalues is complete in $\text{Im } A$. In the case where $\text{Ker } A = 0$ this means that for the operator $A^{-1}: \text{Im } A \to E$ the system of root vectors is dense in its domain of definition (and it turns out to be dense in $E$ if $D(A^{-1}) = \text{Im } A$ is dense). In typical concrete situations, $A^{-1}$ is the initial operator, for example, if $A^{-1} = L$, a differential operator, then $A$ is the integral operator defined by the Green function of the operator $L$. Any selfadjoint (and even any normal) compact operator $A$ on a Hilbert space is complete. It is natural to expect that completeness is preserved under small perturbations of these operators. This expectation is justified if the eigenvalues of the unperturbed operator tend to zero sufficiently rapidly. The first general result in this direction is due to M.V. Keldysh (1951).

**Theorem of Keldysh.** Suppose that the operator $A$ on a Hilbert space $E$ has the form $A = S(1 + R)$, where $S$, $R$ are compact operators, $S$ is selfadjoint and the eigenvalues $\sigma_n \neq 0 (n = 0, 1, 2, \ldots)$ of $S$, taken according to their multiplicities, satisfy the condition

$$\sum_{n=0}^{\infty} \sigma_n^n < \infty$$

(41)

for some $p > 0$. If $\text{Ker } A = 0$ and $\text{Ker } S = 0$, the operator $A$ is complete.

Condition (41) is equivalent to the requirement that $S$ belongs to the class $\mathcal{G}_p$. We shall clarify this from a certain more general point of view.

Let $T$ be any compact operator on a Hilbert space $E$. Then $T^* T$ is a selfadjoint non-negative compact operator. Its positive eigenvalues can be enumerated according to multiplicity in decreasing order: $\tau_1 \geq \tau_2 \geq \cdots$. The numbers $s_1(T) = \sqrt{\tau_1}$, $s_2(T) = \sqrt{\tau_2}$, \ldots are called the singular numbers (s-numbers) of the operator $T$ (clearly $s_1(T) = \|T\|$). It turns out that they coincide with the a-numbers (D. Eh. Alakhverdiev, 1957). The class $\mathcal{G}_p$ for a Hilbert space is therefore determined by the condition

$$\sum_n s_n^n(T) < \infty.$$

For a selfadjoint operator this condition is equivalent to (41).

**Remark 1.** It can be shown that if $\text{Ker } T = 0$ and $\text{Ker } T^* = 0$, the operators $T^* T$ and $T T^*$ are similar (even unitarily similar). Therefore $T^* \in \mathcal{G}_p \Leftrightarrow T \in \mathcal{G}_p$.

---

86 We denote the system of root vectors corresponding to all eigenvalues by $\Sigma(A)$.

87 If $T$ is of finite rank $r$ then we put $\tau_n = 0 (n > r)$. 
Remark 2. For any compact operator $T$ on a Hilbert space we have a Schmidt expansion

$$T = \sum_n s_n(T)(., u_n)v_n,$$

where $(u_n), (v_n)$ are two orthonormal systems. The first consists of the eigenvectors of the operator $A_1 = \sqrt{T^*T}$ while the second is made up of those of the operator $A_2 = \sqrt{TT^*}$. This result is closely connected with polar representations: $T = U_1A_1A_2$, where $U_1: \text{Im} T^* \to \text{Im} T, U_2: \text{Im} T \to \text{Im} T^*$ are invertible isometries.

Moving on to a concise account of the proof of Keldysh theorem, we recall that $\mathcal{S}_p$ is a two-sided (not closed) ideal in the algebra $\mathcal{L}(E)$ of all bounded operators on the Hilbert space. We denote by $P$ the orthoprojection onto the closed linear hull of the system $\Sigma_0(A) = \Sigma(A)$ and put $Q = 1 - P$. Then the operator $\tilde{A} = QAQ = QA$ turns out to be a Volterra operator and therefore its Fredholm resolvent $F(\zeta) = (1 - \zeta \tilde{A})^{-1}$ is an entire function of $\zeta$. Moreover $\tilde{A} \in \mathcal{S}_p$. We can deduce from this that $F(\zeta)$ has growth of order no higher than $p$ and of minimal type as $|\zeta| \to \infty$: $\ln \|F(\zeta)\| = o(|\zeta|^p)$. On the other hand, using the given structure of the operator $A$, we can show that $\|F(\zeta)\|$ is bounded on the pair of angles $\varepsilon \leq |\arg \zeta| \leq \pi - \varepsilon$. Boundedness on the pair of angles $|\arg \zeta| \leq \varepsilon, \pi - \varepsilon \leq |\arg \zeta| \leq \pi$ follows from the Phragmen-Lindelöf principle. Finally, $F(\zeta)$ is bounded on the whole plane. By Liouville's theorem $F(\zeta)$ is constant. Consequently $\tilde{A} = 0$, i.e. $QA = 0$, which establishes the required inclusion $\text{Im} A \subset \text{Im} P = \text{Lin} \Sigma_0(A)$.

**Corollary.** If $S, R$ satisfy the conditions of the theorem, then the operator $B = (1 + R)S$ is complete.

In fact $\text{Ker} (1 + R) = 0$, therefore we have from Fredholm theory that $1 + R$ is an invertible operator; but then $B$ is similar to $A$: $B = (1 + R)A(1 + R)^{-1}$. □

Remark. V.I. Matsaev (1961) established that it is sufficient in Keldysh's theorem to require that the operator $S$ belongs to the two-sided ideal $\mathcal{S}_o$ which is defined by the condition: $\sum n^{-s} s_n(A) < \infty$. He developed a general method for estimating the growth of the resolvent using the distribution function of the $s$-numbers, from which he was able to obtain profound tests for completeness of a system and summability of the corresponding expansion. We note here that if $A, B$ are selfadjoint operators with $B \in \mathcal{S}_o$ and $\text{spec}(A + Bi) \subset \mathbb{R}$, then $A + Bi$ is an operator with separated spectrum and a complete system of spectral subspaces (corresponding to an arbitrarily fine covering of the spectrum), since its resolvent satisfies Levinson's condition (V.I. Matsaev, 1961).

**Example.** Let us consider the differential operator $A$ which is defined on $L^2(a, b)$ ($(a, b)$ a finite interval) by the expression

\[\text{The only non-trivial closed two-sided ideal in } \mathcal{L}(E) \text{ is the ideal of compact operators (theorem of Calkin).}\]
[Image 0x-8 to 403x646]

Chapter 2. Foundations and Methods

\[ l[y] = \left( \frac{1}{i} \frac{d}{ds} \right)^n + \sum_{k=2}^{n} p_k(s) \frac{d^{n-k}}{ds^{n-k}} \]

and boundary conditions

\[ \Phi_j[y] = \sum_{k=0}^{n-1} [a_{jk}y^{(k)}(a) + b_{jk}y^{(k)}(b)] = 0 \quad (1 \leq j \leq n). \]

Under these conditions let the principal part

\[ l_0 = \left( \frac{1}{i} \frac{d}{ds} \right)^n \]

define a selfadjoint operator \( H \) and let both operators \( A, H \) be injective (the latter does not restrict generality). We put \( S = H^{-1} \). This is a compact selfadjoint operator. Its eigenvalues have the form \( \lambda = i\omega^n \), where \( \omega \) runs through the set \( \Omega \) of roots of the characteristic determinant

\[ \Delta(\omega) = \det(\Phi_j[\phi_{n,k}])_{j,k=1}^{n} \]

(the \( \varepsilon_k \) are the \( n \)th roots of unity). The multiplicities of the eigenvalues do not exceed \( n - 1 \). Since \( \Delta(\omega) \) is an entire function of exponential type we have

\[ \sum_{\omega \in \Omega} \frac{1}{|\omega|^{1+\delta}} < \infty \]

for any \( \delta > 0 \). The operator \( S \) therefore satisfies the conditions of Keldysh's theorem for any \( p > 1/n \). We now have

\[ A = \left( 1 + \sum_{k=2}^{n} p_k \frac{d^{n-k}}{ds^{n-k}} S \right) H. \]

But

\[ \sum_{k=2}^{n} p_k \frac{d^{n-k}}{ds^{n-k}} S \]

is a Hilbert-Schmidt integral operator and is therefore compact. Consequently \( A^{-1} = S(1 + R) \), where \( R \) is a compact operator. Applying Keldysh’s theorem we establish that the system of root functions of the operator \( A \) is complete in \( L^2(a, b) \).

Similarly we can obtain completeness theorems for systems of root functions of partial differential operators of elliptic type on bounded domains under boundary conditions, where the principal part of the operator is selfadjoint.

Under conditions which are strictly stronger than those of Keldysh's theorem it is possible to obtain a theorem on the expansion of each vector \( x \in E \) in a series

---

\(^{89}\) We assume that the boundary conditions are linearly independent and the coefficients \( p_n(s) \) are measurable and essentially bounded.

\(^{90}\) The weaker result that \( S \in \mathbb{S}_2 \) follows from the fact that \( S \) is an integral operator with Hilbert-Schmidt kernel.
of root vectors of the operator $A$, which converges after a certain grouping (independent of $x$) of its terms (A.S. Markus, 1962).

In a generalisation of Keldysh's theorem to operators on a Banach space (A.S. Markus, 1966) selfadjointness has a suitable replacement in the property of being quasiconservative.

**Theorem.** Suppose that an operator on a Banach space has the form $A = S(I + R)$, where $S, R$ are compact operators and in addition $S$ is a quasiconservative operator of class $\mathcal{G}_p$ for some $p > 0$. If $\ker A = 0$ and $\ker S = 0$, the operator $A$ is complete.

The eigenvalues of any compact operator can be enumerated in order of decreasing modulus, taking account of multiplicity: $\lambda_1, \lambda_2, \ldots$ (if the set of eigenvalues is finite we complete the infinite "tail" with zeros). If the series $\sum_n \lambda_n$ converges unconditionally (and so also absolutely since it is a series of numbers) then its sum is called the *spectral trace* of the operator $A$. A natural and important problem is to determine conditions under which the spectral trace coincides with the tensor trace (clearly such conditions must ensure the existence of both traces). The first advance in this direction (not counting the classical theorem for the finite dimensional case) is due to V.B. Lidskij (1959), who showed that for any operator of class $\mathcal{G}_1$ on a separable Hilbert space $E$ (in other words, the operator is nuclear since $E$ is a Hilbert space) both the spectral trace and the matrix trace with respect to any orthonormal basis exist and they are equal. By the same token the tensor trace is equal to the spectral trace in the given case. In further investigations (A.S. Markus – V.I. Matsaev (1971), König (1986)) this result was extended to operators of class $\mathcal{G}_1$ on any Banach space.

One of the important applications of Lidskij's theorem is to provide a necessary condition for completeness of the system $\Sigma(A)$ of root vectors of a compact operator $A$ on a separable Hilbert space, whose imaginary component $\text{im } A$ is nuclear:

$$\sum_{n=1}^{\infty} \text{im } \lambda_n = \text{tr}(\text{im } A). \tag{42}$$

For the proof it is sufficient to orthogonalise the system $\Sigma(A)$. The diagonal elements of the matrix of the operator $\text{im } A$ with respect to the resulting orthonormal basis are equal to $\text{im } \lambda_n (n = 1, 2, \ldots)$.

For a compact dissipative operator $A$ on a Hilbert space, whose real component is nuclear, the "trace equation", rewritten in the form

$$\sum_{n=1}^{\infty} \text{re } \lambda_n = \text{tr}(\text{re } A), \tag{42'}$$

---

91 A fundamental result in this situation is that the matrix trace of a Volterra operator of class $\mathcal{G}_1$ is equal to zero.

92 If $A$ is a bounded operator ($D(A) = E$) then

$$\text{re } A = \frac{1}{2 \text{det } 2} (A + A^*), \quad \text{im } A = \frac{1}{2 \text{det } 2} (A - A^*).$$
is not only necessary but also sufficient for completeness of the system $\Sigma(A)$ (M.S. Livshits, 1954). In fact, for the invariant subspace $E_1 = \text{Lin} \Sigma(A)$ the trace of the operator $\text{re}(A|_{E_1})$ is given by (42'). If $E_1 \neq E$ then $E_0 \equiv E_1 \neq 0$ and then for the orthoproduction $P$ on $E_0$ we will have $\text{tr}(P \cdot \text{re}(A|_{E_0}) \cdot P) = 0$; hence $\text{re}(A|_{E_0}) = 0$ since $\text{re} A \leq 0$. Consequently $(1/i)A|_{E_0}$ is selfadjoint ($E_0$ is invariant for $A^*$). Since $A|_{E_0}$ is compact it must have an eigenvector – a contradiction.

It is clear from what has just been discussed that the inequality

$$\sum_{n=1}^{\infty} \text{re} \lambda_n \geq \text{tr}(\text{re} A)$$

holds for any compact dissipative operator with nuclear real component. The system of root vectors of a nuclear dissipative operator is complete (V.B. Lidskij, 1959) since $\sum_n \lambda_n = \text{tr} A$ and $\sum_n \overline{\lambda_n} = \text{tr} A^*$.

In conclusion we note that completeness of a compact operator is not in general inherited by its restrictions93: any Volterra operator on a Banach space is the restriction of a complete compact operator acting on some larger Banach space (N.K. Nikol'skij, 1969).

4.14. Triangular Decompositions. The theorem on the existence of a (non-trivial) closed invariant subspace for any compact operator $A$ on a complex LCS $E$ ($\dim E > 1$), with the help of Zorn's lemma, leads to the existence of a chain of such subspaces which is maximal with respect to inclusion. Any such chain begins with zero and terminates with the space $E$. In the finite-dimensional case it has the form $0 \subset L_0 \subset L_1 \subset \cdots \subset L_n = E$, where $\dim I_{k} = k$ ($0 \leq k \leq n = \dim E$). Any system of vectors $\{e_k\}_1^n$ chosen so that $e_k \in L_k \setminus L_{k-1}$ ($1 \leq k \leq n$) is a basis in $E$ with respect to which the matrix of the operator is upper triangular.

In the infinite-dimensional case the chain $Z$ is not in general discrete (for the order topology) and therefore it cannot generate a basis for a triangular decomposition. The latter has to be connected directly with the chain $Z$. For each non-zero subspace $M \in Z$ we consider the closure $M_\infty$ of the union of all those $L \in Z$ which are contained in $M$ but are distinct from it. Then $M_\infty \in Z$ since the chain $Z$ is maximal and clearly $M_\infty \subset M$. If $M_\infty \neq M$, the factor space $M/M_\infty$ is one-dimensional. Those $M$ for which this holds are called points of discontinuity of the chain $Z$. In the absence of points of discontinuity the chain $Z$ is said to be continuous.

If $M$ is a point of discontinuity, the factor operator $A_M$ can be identified with some scalar $\lambda_M$. Ringrose (1962) proved the following theorem (for a Banach space $E$).

**Theorem.** The set of all non-zero values $\lambda_M$ corresponding to all possible points of discontinuity coincides with $\text{spec} A \setminus \{0\}$. For each $\lambda \in \text{spec} A \setminus \{0\}$ the number

93 If an operator is complete along with all of its restrictions, then we say that it allows spectral synthesis. The problem of spectral synthesis goes back to a work of Wermer (1952) which is concerned with normal operators. The general situation was studied by A.S. Markus (1970).
of points of discontinuity for which $\lambda_M = \lambda$ is equal to the dimension of the corresponding root subspace.

**Corollary.** In order that the operator $A$ be a Volterra operator it is necessary and sufficient that $\lambda_M = 0$ for all points of discontinuity of the chain $Z$.

In particular, if the chain $Z$ is continuous, $A$ must be a Volterra operator. Conversely, if $A$ is a dissipative Volterra operator on a Hilbert space $E$ and $\text{Ker } A = 0$, then all maximal chains of its closed invariant subspaces are continuous (M.S. Brodskij, 1969).

A continuous maximal chain $Z$ for any Volterra operator $A$ on a Hilbert space $E$ results from inessential extension. In this construction a certain Hilbert space $G$ (dim $G$ depends on $A$) is adjoined orthogonally to $E$ and we take the operator $A$ on $H = E \oplus G$ such that $A|_E = A$, $A|_G = 0$; $A$ has the required chain.

A Volterra operator $A$ is said to be **unicellular** if the maximal chain of closed invariant subspaces is unique. This term originates from the simplest finite-dimensional example - a single Jordan cell. A non-trivial example is the integration operator $I$ on $L^p(0, 1)$ ($1 \leq p \leq \infty$). In fact, it follows from the Titchmarsh convolution theorem and the completeness of the system of powers that any closed invariant subspace of the operator $I$ must consist of all functions which vanish almost everywhere on a fixed interval $(0, a)$ ($0 < a < 1$). In exactly the same way we can establish the unicellularity of all integral operators of fractional order:

$$(I_\alpha \varphi)(s) = \frac{1}{\Gamma(\alpha)} \int_0^s (s - t)^{\alpha - 1} \varphi(t) dt \quad (0 < \alpha \leq 1).$$

It is interesting that no two of these operators on $L^2(0, 1)$ are equivalent (I.Ts. Gokhberg – M.G. Krejn, 1967), a result which contrasts sharply with the equivalence of all unicellular operators on a finite-dimensional space.

In a Hilbert space $E$ instead of closed subspaces we can consider the corresponding orthoprojections. Corresponding to the subspace inclusion $L_1 \subset L_2$ we have the inequality $P_1 \leq P_2$ between the respective projections and therefore the chain of subspaces transforms into a chain of orthoprojections. Suppose we are given an arbitrary strongly closed chain $W$ of orthoprojections and a function $F: W \to \mathcal{B}(E)$. We can define the **Brodskij integral**

$$A = \int_W F(P) dP$$

as the uniform limit of the integral sums

$$\sum_{k=1}^n F(Q_k)(P_k - P_{k-1}).$$

---

94 This is equivalent to pairwise comparability of the closed invariant subspaces.

95 A further example: the weighted shift operator $(\xi_n \xi_0)_{n \to (\xi_n \xi_{n-1})_0} (\xi_{-1} = 0)$ on $l^2$ under specific conditions on the weight sequence $(\alpha_n)$ (N.K. Nikolskij, 1967).
directed by the partitions $P_0 < P_1 \cdots < P_n$, where all the $P_k$, $Q_k$ belong to the chain $W$ and $P_k-1 \leq Q_k \leq P_k (1 \leq k \leq n)$. It is clear that not every function $F$ is integrable in this sense.

If $A$ is a Volterra operator, $Z$ is a continuous maximal chain of closed invariant subspaces and $W$ is the corresponding chain of orthoprojections, then we have the triangular decomposition (M.S. Brodskij, 1961)

$$A = 2i \int_W P(\text{im } A) \, dP$$

(the existence of the integral is guaranteed). From this we extract a rich supply of information on relationships concerning properties of the real and imaginary components of a Volterra operator; for example there is the theorem of Sakhnovich that, if the imaginary component of a Volterra operator is a Hilbert-Schmidt operator, then so also is its real component; we also have the more general theorem of Matsaev: if $A$ is a Volterra operator and $\text{im } A \in \mathfrak{S}_p$, then $\text{re } A \in \mathfrak{S}_p (1 < p < \infty)$ and moreover $N_p(\text{re } A) \leq \gamma_p N_p(\text{im } A)$, where the constant $\gamma_p$ depends only on $p$. In the same direction for $p = 1$ (i.e. for $\text{im } A$ nuclear) I. Ts. Gokhberg and M.G. Krejn (1967) established the deep spectral density theorem for the operator $\text{re } A$:

$$\lim_{R \to \infty} \frac{n_+(R)}{R} = \frac{1}{\pi} \int_W N_1(\text{im } A \cdot dP),$$

where $n_+(R)$ is the number of eigenvalues (counted according to multiplicity) of the operator $\text{re } A$ in the semiaxis $[R^{-1}, \infty)$.

### 4.15. Functional Models

A model of a linear operator $A$ acting on an LTS $E$ is a commutative diagram

$$
\begin{array}{ccc}
E & \xrightarrow{A} & E \\
\downarrow V & & \downarrow V \\
G & \xrightarrow{r} & G
\end{array}
$$

in which $B$ is a linear operator on some LTS $G$ and $V$ is a topological isomorphism (an operator intertwining) $A$ with $B$. Clearly the operators $A$ and $B$ are equivalent (similar) in this situation. Frequently we just call the operator $B$ a model of the operator $A$. If $E$ and $G$ are Hilbert spaces and $V$ maps $E$ isometrically onto $G$, we use the term unitary equivalence. On replacing an operator by its model all its linear-topological properties and its essential nature are preserved (for example, the spectrum, eigenvalues, approximate spectrum). Moreover a

---

96 Here (as in certain other contexts) the term 'operator' is used in spite of the fact that $V$ acts from one space into another. As in the theory of representations it is convenient to define the term 'intertwining operator' in a more general way, namely to require only continuity of the operator $V$ but not necessarily continuous invertibility. In this situation we can speak of a quasimodel.
concrete model of an operator may prove simpler to study. For example, if it is
a functional model (i.e. $G$ is a function space), some analytic techniques can be
applied to it. By way of illustration let us consider a Volterra operator $A$ on a
Hilbert space with one-dimensional imaginary component\footnote{Any operator with one-dimensional imaginary component is dissipative to within a factor $\pm i$.} and suppose that
\( iA \) is completely non-selfadjoint (it has no selfadjoint parts). It follows from results
of M.S. Livshits (1946) that $A$ is unitarily equivalent to an operator $\pm iI_s$, where
$I_s$ is the integration operator\footnote{The modification $I \mapsto I'$ is occasioned by the desire to work with an upper- (rather than a lower-) triangular decomposition. In other respects it is immaterial.} on $L^2(0, l)$ from $s$ up to $l$ ($0 < l < \infty$). The
unicellularity of the operator $I_s$ (which can be established by analytic means)
therefore implies that of the operator $A$. However the last result can also be
obtained directly from the triangular decomposition of the operator $A$ (M.S.
Brodskij, 1965)\footnote{And Titchmarsh convolution theorem can be deduced from the fact that the operator $I$ is
unicellular (Kalisch, 1962).}. In any event the very fact that the operator $I_s$ is unicellular
entitles us to fix it in the role of a standard triangular model which is not subject
to further decomposition just as with a Jordan cell in the finite-dimensional case.

The more general triangular model

\[
(A\varphi)(s) = \alpha(s)\varphi(s) + i \int_s^t \varphi(t) \, dt \quad (0 \leq s \leq l)
\]

generated by the real function $\alpha \in L^\infty(0, l)$ is no longer a Volterra operator: the
spectrum of the operator $A$ coincides with the set of essential values of the
function $\alpha$ (the term $\alpha(s)\varphi(s)$ is analogous to the diagonal part of a matrix). Any
completely non-selfadjoint bounded operator $A$ with one-dimensional imaginary
component and real spectrum is unitarily equivalent to an operator $A$ defined
by some non-decreasing function $\alpha$ and choice of sign in front of the integral. For
the proof of this theorem M.S. Livshits introduced a powerful unitary invariant, the
characteristic function

\[
w_A(\lambda) = 1 - i(R_\lambda(A)e, e)j,
\]

where the channel vector $e$ and the number $j = \pm 1$ are defined by the equation
$im A = \frac{1}{2}j(., e)e$. It turned out that for the given class of operators coincidence
of characteristic functions implies unitary equivalence. But for an operator $A$ its
characteristic function can be calculated immediately; in fact with $j = 1$ (to be
specific)

\[
w_A(\lambda) = \exp \left( i \int_0^t \frac{dt}{\lambda - \alpha(t)} \right). \tag{43}
\]

Moreover for the operator $A$ under consideration (with $j = 1$) the characteristic
function maps the upper half-plane $\text{im} \lambda > 0$ conformally onto the domain
$|w| > 1$ and is regular at infinity. It can therefore be represented in the form (43)
for some $\alpha(t)$. But then $A$ turns out to be unitarily equivalent to the corresponding
operator $A$.\"
Where there is a non-real spectrum (necessarily no more than countable and consisting of normal eigenvalues), a 'discrete' triangular model is adjoined orthogonally to the previous 'continuous' model; this is defined on $l^2$ by the formulae

$$\eta_k = \lambda_k \xi_k + i \sum_{n=k+1}^{\infty} \xi_n \beta_n \beta_n + i \int_0^1 \varphi(t) \beta_k \, dt \quad (k = 1, 2, 3, \ldots),$$

where $(\lambda_k)^\infty$ is the sequence of eigenvalues and $\beta_k = \text{im} \lambda_k$. Correspondingly, the Blaschke product

$$\prod_{k=1}^{\infty} \frac{\lambda - \lambda_k}{\lambda - \bar{\lambda}_k}$$

appears in (43). Developing this approach, M.S. Livshits (1954) constructed a triangular model for a completely non-selfadjoint bounded operator $A$ with finite-dimensional imaginary component. In this case the characteristic function turns out to be a matrix of order $r = \text{rk}(\text{im} A)$ and the role of $j$ is taken by a certain Hermitian matrix $J$ such that $J^2 = 1$. The characteristic matrix-function $W_A(\lambda)$ (im $\lambda > 0$) is $J$-expanding, i.e. $W_A^*(\lambda)JW_A(\lambda) \geq I$, and moreover on the real axis outside of the spectrum it is $J$-unitary: $W_A^*(\lambda)JW_A(\lambda) = I$. The corresponding analogue of formula (43) (with products of Blaschke type) was obtained by V.P. Potapov (1950). Passing to the limit as $r \to \infty$ we can extend the triangular model further to the case where the imaginary component is nuclear.

It is impossible to get rid of the condition of being completely non-selfadjoint, since the characteristic function does not change if an arbitrary bounded self-adjoint operator is adjoined orthogonally to $A$. However this clearly does not matter from the point of view of classification theory. We can say that for operators whose imaginary components are nuclear there is a valuable analogue of Jordan's theorem.

Incidentally, although the spectral resolution of a selfadjoint operator

$$A = \int_{\mathbb{R}} \lambda \, d\mathcal{E}_\lambda$$

reveals its structure sufficiently deeply, it is nevertheless useful to construct in addition a functional model for it. Its structure depends on the multiplicity of the spectrum which we will not define in the general case; we restrict ourselves to the case of a simple spectrum, i.e. where there exists a vector $x$ such that the system $(\mathcal{E}_\lambda x)_{\lambda \in \mathbb{R}}$ is complete. Then, putting $\sigma(\lambda) = (\mathcal{E}_\lambda x, x)$, we can show without difficulty that $A$ is unitarily equivalent to the operator of multiplication by $\lambda$ on the space $L_2^x$.

Now let $A$ be a contraction on a Hilbert space $E$. A natural measure of the non-unitariness of an operator $A$ is the pair of non-negative selfadjoint operators

100 To be precise it is assumed further that the non-real spectrum is infinite.

101 A vector $x$ with the stated property is said to be generating or cyclic.
$1 - A^*A$ and $1 - AA^*$ (just as the imaginary component of an operator is a measure of its non-selfadjointness). Let us put

$$
\Delta_r = \sqrt{(1 - A^*A)}, \quad \Delta_l = \sqrt{(1 - AA^*)}.
$$

The characteristic operator-function\(^{102}\)

$$
W(\lambda) = (-A + \lambda \Delta_r (1 - \lambda A^*)^{-1} \Delta_r) |_{\text{Im} \Delta_r}
$$
is introduced in this context (Yu.L. Shmul'yan, 1953). The values of this function are contracting homomorphisms from $\text{Im} \Delta_r$ into $\text{Im} \Delta_l$. It is defined and analytic on the domain $(\text{reg} A^*)^{-1}$, in particular on the disk $|\lambda| < 1$, where it is usually considered. If the characteristic functions of completely non-unitary contractions coincide, then these contractions are unitarily equivalent (A.V. Shtraus, 1960). Szőkefalvi-Nagy and Foias (1963) found a natural (in terms of characteristic functions) functional model for an arbitrary completely non-unitary contraction $A$. Of necessity it is rather complicated, since any contraction is the orthogonal sum of a completely non-unitary contraction and a unitary operator and any bounded operator becomes a contraction on normalisation.

The construction of the model is significantly simplified if the sequence of powers $(A^n)_0^\infty$ converges strongly to zero (although even in this case getting information depends essentially on the measure of non-unitariness, more precisely on the rank of the operator $\Delta_r$). In order to describe the model operator in this situation we consider an arbitrary Hilbert space $\mathcal{G}$ and introduce the Hardy class $H^2(\mathcal{G})$ of analytic vector-functions

$$
u(\lambda) = \sum_{n=0}^\infty u_n \lambda^n \quad (|\lambda| < 1, \ u_n \in \mathcal{G} \ (n = 0, 1, 2, \ldots)),
$$
which satisfy the condition

$$
\|\nu\|^2 = \sum_{n=0}^\infty \|u_n\|^2 < \infty.
$$

This is a Hilbert space with natural scalar product

$$
(u, v) = \sum_{n=0}^\infty (u_n, v_n).
$$

The shift operator\(^{103}\) $(Tu)(\lambda) = \lambda u(\lambda)$ acts on it; it is clearly isometric (but not unitary: $\text{Im} T = \{u: u(0) = 0\}$). The adjoint operator

$$
(T^*v)(\lambda) = \frac{v(\lambda)}{\lambda} = \sum_{n=0}^\infty v_{n+1} \lambda^n \quad \left(v(\lambda) = \sum_{n=0}^\infty u_n \lambda^n\right)
$$
is a contraction (and the sequence of its powers converges strongly to zero).

---

\(^{102}\) This construction arose in another way in joint works of Szőkefalvi-Nagy and Foias (1962) and for finite-dimensional $\Delta_l$, $\Delta_r$ it is equivalent to the definition of M.S. Livshits (1950). A.V. Shtraus (1959) extended the concept of the characteristic function to unbounded operators.

\(^{103}\) Strictly speaking, the shift is effected on the sequence of Taylor coefficients of the function $u(\lambda)$. 
we now take $G = \text{Im} \Delta$, then $T^*$ turns out to be a quasimodel for $A$. The intertwining operator $V: E \to H^2(G)$ is defined by the formula

$$(Vx)(\lambda) = \sum_{n=0}^{\infty} \lambda^n (A^nx).$$

In fact

$$\|Vx\|^2 = \sum_{n=0}^{\infty} (\Delta x, A^{n+1}x) = \sum_{n=0}^{\infty} ((1 - A^*A)A^n x, A^n x)$$

$$= \sum_{n=0}^{\infty} \{ \|A^n x\|^2 - \|A^{n+1} x\|^2 \} = \|x\|^2,$$

i.e. $V$ maps $E$ isometrically into $H^2(G)$. In addition it is clear that $T^*V = VA$. Therefore the subspace $L = \text{Im} V \subset H^2(G)$ is invariant for $T^*$ and $T^*|_L$ is unitarily equivalent to the operator $A$.

If $\text{rk} \Delta = 1$, the model operator is the adjoint of a shift on the scalar Hardy class $H^2$. It is possible to describe all closed invariant subspaces for this operator (and hence also those for the initial operator $A$) (Beurling, 1949). Each such (non-zero) subspace has the form

$$\theta H^2 = \{ \phi: \phi = \theta \psi (\psi \in H^2) \},$$

where $\theta$ is a fixed function in $H^2$ whose boundary values on the unit circle have modulus one almost everywhere (an interior function).

The general approach to models of Szökefalvi-Nagy and Foias is based on the study of the so-called dilatations $\tilde{A}$ of the operator. This term is applied to any operator $\tilde{A}$ which acts on some Hilbert space $E \supset E_0 \supset \ldots$ in which $E$ is embedded via the first summand. It has the form

$$A^n = P \tilde{A}^n \quad (n = 1, 2, 3, \ldots),$$

where $P$ is an orthoprojection on $E$ with $\text{Im} P = E$.

**Theorem of Szökefalvi-Nagy.** For any contraction $A$ on a Hilbert space there exists a unitary dilatation.

The original proof (1959) of this important theorem was connected with the trigonometric moment problem and the theory developed by M.A. Najmark (1940) of selfadjoint extensions of operators out of their basic space. There is, however, a more direct proof (Szökefalvi-Nagy and Foias, 1967). First of all an isometric dilatation is constructed on $E = E \oplus E \oplus \ldots$ in which $E$ is embedded via the first summand. It has the form

---

104 Figuring in this theory are non-orthogonal spectral resolutions of symmetric operators

$$A = \int_{-\infty}^{\infty} \lambda d \mathfrak{F}_\lambda,$$

where $\mathfrak{F}_\lambda = P_{E_\lambda}$, $E_\lambda$ being an orthogonal resolution of the identity in a larger Hilbert space $\tilde{E}$ and $P$ an orthoprojection on $\tilde{E}$ onto $E$. 

---
Now it is possible to extend an isometric dilatation to a unitary one. In fact, thanks to the Wold decomposition, we can suppose that the initial dilatation is a one-sided shift and then the corresponding two-sided shift will be its unitary extension.

The theory of characteristic functions, functional models and dilatations found deep applications in mathematical physics. In particular, in quantum-mechanical scattering problems the characteristic matrix-function coincided with the so-called $S$-matrix of Heisenberg (M.S. Livshits, 1956). This result is of a sufficiently general nature (Lax-Phillips, 1964; V.M. Adamyan – D.Z. Arov, 1965). It was used for the investigation of analytic properties of the $S$-matrix (B.S. Pavlov – L.D. Faddeev, 1972) in scattering theory for automorphic functions, which Lax and Phillips (1976) then developed in detail. The theory of dilatations proved to be an effective tool for investigating the so-called Redge singularity (B.S. Pavlov, 1970).

From the physical point of view, non-selfadjointness or non-unitariness of the basic operators which describe the dynamics of physical systems reflects processes of dissipation (of energy or other substance) caused by non-closedness (openness) of the systems, i.e. by their relations with the external world. Starting from these ideas M.S. Livshits (1963) associated with each open system a mathematical object called an operator node. This consists of a pair of Hilbert spaces $H, E$ ($H$ is the space of internal states of the system, $E$ is the space of channels of communication of the system with the outside world) and a set of three continuous homomorphisms: a dynamical operator $A: H \to H$, a channel operator $J: E \to E$ ($J^* = J, J^2 = 1$) and a communication operator $\Gamma: E \to H$. It is assumed in addition that the relation $\text{im } A = \mathcal{T}J \Gamma^*$ is satisfied. Unitary equivalence is defined naturally in the category of operator nodes. The characteristic operator-function

$$W(\lambda) = 1 - iJ \Gamma^* \mathcal{R}_A(A) \Gamma$$

is a basic unitary invariant of an operator node. If an open system is decomposed into a chain of subsystems the characteristic function becomes factorised. Conversely, therefore, if we factorise the characteristic function into factors corresponding to elementary (in the physical sense) systems, we then obtain a model for the initial system in the form of a chain of elementary systems. Many of the concrete situations can be included in this scheme (M.S. Livshits, 1966).

### 4.16. Indefinite Metric

The structure introduced on a Hilbert space $E$ by a selfadjoint operator $J$ such that $J^2 = 1$ ($J \neq \pm 1$) is called an indefinite metric,

---

105 From external applications of this theory we note that of necessary and sufficient conditions for the existence of Riesz and Bari bases composed of root vectors of a contraction on a Hilbert space (N.K. Nikolskij - B.S. Pavlov, 1970).

106 Among the channels of communication are entrance and exit channels. Successive unions of systems are effected by connecting the exits of one system to the entrances of another. This procedure can be formalised in terms of what has been described.
or J-metric, since it allows us to consider along with the usual distance the pseudodistance

\[ d_J(x, y) = \sqrt{(J(x - y), x - y)} \]

(with values in \( \mathbb{R} \cup i\mathbb{R} \)) defined by the J-scalar product \( \langle u, v \rangle = \langle Ju, v \rangle \). The Lorentz metric in four-dimensional space, which is used in the special theory of relativity, is an example of this. Indefinite metrics on finite-dimensional spaces had already been studied by Frobenius (at the end of the 19th century). The first work dealing with operators on an infinite-dimensional space with an indefinite metric is due to L.S. Pontryagin (1944), further development being associated first of all with investigations of M.G. Krejn and I.S. Iokhvidov. At the present time the theory of linear operators on spaces with an indefinite metric represents by itself an extensive branch of functional analysis with outlets into numerous applications (theory of oscillations, quantum physics etc.).

Clearly, we can represent the metric operator as the difference of two mutually orthogonal (non-zero) orthoprojections which form a decomposition of the identity:

\[ J = P_+ - P_-; \quad P_+ + P_- = 1; \quad P_+P_- = P_-P_+ = 0. \]

Correspondingly, any \( x \in E \) can be expressed as the sum of mutually orthogonal vectors \( x = P_+x, x = P_-x; \) moreover

\[ Jx = x_+ - x_-, \quad \langle x, x \rangle = \|x_+\|^2 - \|x_-\|^2. \]

The index \( \kappa = \text{rk } P_- \) is the measure of indefiniteness of the space. It was assumed in the work of L.S. Pontryagin that \( \kappa \) is finite and many results were obtained initially under this condition.

Remark. Suppose that on a complex linear space \( E \) there is given a Hermitian symmetric bilinear functional \( \langle u, v \rangle \) which is non-degenerate in the sense that for each \( u \neq 0 \) there exists \( v \) such that \( \langle u, v \rangle \neq 0 \). Let us assume that for the restriction of this functional to all possible finite-dimensional subspaces the maximum of the negative indices of inertia is equal to \( \kappa < \infty \). Then clearly this maximum is attained at some \( \kappa \)-dimensional subspace \( E_- (\langle u, u \rangle < 0 \text{ for all } u \in E_-, u \neq 0) \). Let us denote by \( E_+ \) the space of vectors which are J-orthogonal to \( E_- \). Then \( \langle v, v \rangle > 0 \) for all \( v \in E_+, v \neq 0 \) (i.e. \( E_+ \) is pre-Hilbert under the J-metric) and \( E = E_- + E_+ \). Starting with the decomposition \( x = u + v \) (\( u \in E_-, v \in E_+ \)) we introduce a scalar product on \( E(x_1, x_2) = -\langle u_1, u_2 \rangle + \langle v_1, v_2 \rangle \), thereby making \( E \) into a pre-Hilbert space. Completeness of the space \( E \) is equivalent to completeness of the space \( E_+ \) since \( \kappa < \infty \). If we have completeness, \( E \) turns out to be a space with indefinite metric: \( J = P_+ - P_- \), where \( P_\pm \) are the projections associated with the decomposition \( E = E_- + E_+ \).

\[ ^7 \text{If } E \text{ is finite-dimensional, } \kappa \text{ is the negative index of inertia of the form } \langle x, x \rangle. \]

\[ ^{108} \text{It does not depend on the initial choice of the space } E_- \.]
In what follows, $E$ is a Hilbert space with indefinite metric; the notation introduced earlier remains fixed.

The general form of a continuous linear functional on $E$, i.e. $f(x) = (x, y)$, is easily reformulated in $J$-terminology: if $z = Jy$, then $f(x) = \langle x, z \rangle$.

A subspace $L \subset E$ is said to be positive (non-negative) if $\langle x, x \rangle > 0$ for all $x \in L$, $x \neq 0$ ($\langle x, x \rangle \geq 0$ for all $x \in L$). The following terms are introduced in a similar way: negative (non-positive) subspace. A subspace $L$ is said to be neutral if $\langle x, x \rangle = 0$ for all $x \in L$. From the geometrical point of view $\langle x, x \rangle = 0$ is the equation of a second order cone $\mathcal{C}^{109}$; the positive vectors lie 'inside' the cone, the negative ones 'outside' it.

The subspace $E_+ = \text{Im} \ P_+$ is non-negative and maximal (with respect to inclusion) among the non-negative subspaces. Phillips (1959) and Yu. P. Ginzburg (1961) obtained the following important characterisation of maximal non-negative subspaces.

**Theorem.** In order that $L$ be a maximal non-negative subspace it is necessary and sufficient that there exist a contracting linear operator $K: E_+ \to E_-$ ($E_- = \text{Im} \ P_-$) such that $L = \text{Im} (1 + K)$.

Let $A$ be a linear and (for simplicity) bounded operator on $E$. The operator $A' = JA^*J$ is called the $J$-adjoint of $A$ on the grounds that the identity $\langle Ax, y \rangle = \langle x, A \rangle$ is satisfied by it (and only by it). The formal properties of the adjoint are preserved in the $J$-adjoint: $A'' = A'$; $(aA + \beta B)' = \overline{\alpha} A' + \overline{\beta} B'$; $(AB)' = B'A'$; $1' = 1$.

If $A' = A$ we say that $A$ is $J$-selfadjoint. If $A$ is invertible and $A^{-1} = A'$ then $A$ is said to be $J$-unitary.\(^{111}\)

The study of $J$-selfadjoint operators reduces to that of $J$-unitary operators with the help of the Cayley transform. For the sake of brevity therefore, unless otherwise stated, the assertions made in the sketch below relate only to a $J$-unitary operator $A$. First of all we note that it follows from the equality $A^{-1} = J^{-1}A^*J$ that the spectrum of the operator $A$ is symmetric with respect to the unit circle, i.e. if $\lambda \in \text{spec} \ A$ then $\lambda^* \equiv (\overline{\lambda})^{-1} \in \text{spec} \ A$. The spectrum cannot be unitary even in the finite-dimensional case. In general, a meaningful spectral theory for $J$-unitary operators requires that the index $\kappa$ is finite or at least the operator $P_- \ AP_+$ is compact. In this case $A = U + V$ where $U$ is unitary and $V$ is compact. Consequently, the non-unitary spectrum of the operator $A$ is no more than countable (if it is infinite, it condenses on the unit circle) and consists of normal eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots (|\lambda_k| < 1 \ (k = 1, 2, \ldots))$. The corresponding maximal root subspaces $W_1, W_2, W_3, \ldots$ are neutral; each $W_k (W_k^*)$ is

---

109 Here the word 'one' is used in the sense of elementary analytic (if preferred - algebraic) geometry.
110 Because of its geometric character we call this an angular operator.
111 This is equivalent to the identity $\langle Ax, Ay \rangle = \langle x, y \rangle$ in combination with the requirement of surjectivity; by itself the identity, which is known as $J$-isometry, is equivalent to the equality $A^*JA = J$. The inequality $A^*JA \leq J$ (i.e $\langle Ax, Ay \rangle \leq \langle x, y \rangle$) defines the $J$-contractions. We can define $J$-dilatations similarly.
J-orthogonal to all $W_j, W_j^*$ except $W_k^*(W_k)$; the subspaces $W_k, W_k^*$ ($k = 1, 2, \ldots$) are in duality with respect to the $J$-scalar product (consequently, $\dim W_k^* = \dim W_k$).

Suppose that the index $\kappa$ is finite. Then the dimension of the spectral subspace $L_0$ corresponding to the part of the spectrum which lies in the disk $|\lambda| < 1$ does not exceed $\kappa$ and is equal to the dimension of the spectral subspace $L_\infty$ corresponding to the symmetric part of the spectrum. The direct sum $L_k = L_0 + L_\infty$ is called the hyperbolic subspace of the operator $A$. The spectral subspace $L_1$ corresponding to the unitary spectrum is $J$-orthogonal to $L_k$ and $L_1 + L_k = E$.

By the same token (for finite $\kappa$) we can restrict attention to the case of unitary spectrum in the spectral analysis of $J$-unitary operators. In this connection substantial use is made of the

**Theorem of Pontryagin.** If the index $\kappa$ is finite, any $J$-unitary operator $A$ has invariant maximal non-positive and non-negative subspaces.

The dimension of an invariant maximal non-positive subspace $M$ is equal to $\kappa$. It can be chosen in such a way that the spectrum of the operator $A|_M$ lies outside the disk $|\lambda| < 1$. Let $p(\lambda)$ be the characteristic polynomial of the operator $A|_M$ (so that $p(A|_M) = 0$). Then $\text{Im } p(A) \subset M^{(1)}$. But $M^{(1)}$ is a non-negative subspace. Consequently $\langle p(A)y, p(A)y \rangle \geq 0$ ($y \in E$). Putting $q(\lambda) = p(\lambda)p(\lambda^{-1})$, we obtain $\langle q(A)y, y \rangle \geq 0$ and also $q$ is a Hermitian-symmetric (real on the unit circle) polynomial in $\lambda$ and $\lambda^{-1}$. This procedure for achieving definiteness (I.S. Iokhvidov, 1950) served as the key to the spectral theorem which is finally established for all specified $J$-unitary operators with unitary spectrum (Langer, 1967). Although the procedure depends on the choice of subspace $M$, the final spectral resolution is unique. The main complication in the proof of the spectral theorem is associated with the roots of the polynomial $q$ which lie on the circle $|\lambda| = 1$. Near to these critical points, the resolvent has power order of growth higher than the first.

M.G. Krejn and Langer (1964) applied the indefinite metric to investigate the quadratic equation

$$X^2 + BX + C = 0$$

in the algebra $\mathfrak{L}(H)$ of operators on a Hilbert space $H$. Here the coefficient $B$ is assumed to be selfadjoint while $C$ is compact, selfadjoint and non-negative. The operators

\[ X \in \mathfrak{L}(H) \]

are also under the condition that the operator $P_+AP_+$ is compact (M.G. Krejn, 1950). It is interesting to note that M.G. Krejn's proof is based on the fixed point principle.

A $J$-selfadjoint operator was considered in the work of L.S. Pontryagin. The transition to a $J$-unitary operator was effected by I.S. Iokhvidov (1949) using the Cayley transform.

For any subspace $L$ the subspace $L^{(1)} = \{x: \langle x, y \rangle = 0, (y \in L)\}$ can be called its $J$-orthogonal quasicomplement. In general $L \cap L^{(1)} \neq 0$.

Without restriction of the type to compactness or finiteness of the index.

$J$-selfadjoint operators were considered in the work of Langer (the finite index case is due to M.G. Krejn and Langer (1963)).
on the space \( E = H \oplus H \) were considered. Since \( A \) is \( J \)-selfadjoint and \( P \cdot A \cdot P = \sqrt{C} \) is compact, there exists a maximal non-negative subspace which is invariant for \( A \). If \( K \) is the corresponding singular operator, then \( X = K \sqrt{C} \) turns out to be a root of the equation (44). Also \( X^* X \leq C \) since \( \| K \| \leq 1 \).

We note that equation (44) is naturally associated with the equation of linear oscillations

\[
T \ddot{u} + R \dot{u} + Vu = 0
\]

(45)

Here \( u = u(t) \) is a vector-function with values in a Hilbert space \( H \) and \( T, R, V \) are selfadjoint (generally unbounded) operators with \( T > 0, R \geq 0 \) and \( V \gg 0 \) (the forms generated by \( T \) and \( V \) are the kinetic and potential energies respectively, \( R \) is a damping operator). Under the substitution \( v = \sqrt{V} u \), equation (45) becomes

\[
C \ddot{v} + B \dot{v} + v = 0
\]

(46)

where \( C = V^{-1/2} TV^{1/2}, B = V^{-1/2} RV^{1/2} \). The operators \( C \) and \( B \) are usually bounded in applications and after closure they satisfy the conditions formulated above. Exponential solutions \( v = v_0 e^{\mu t} \) arise as eigenvectors of the quadratic pencil of operators \( S(\mu) = C \mu^2 + B \mu + 1, S(\mu) v_0 = 0 \). If \( X \) is an operator root of equation (44), \( \lambda \) is any of its eigenvalues \((\lambda \neq 0)\) and \( v_0 \) is a corresponding eigenvector, then \( v_0 \) is an eigenvector of the pencil \( S \) for the eigenvalue \( \mu = \lambda^{-1} \).

A root \( X \) gives rise to another root \( Y = -X^* - B \), which is also a source of eigenvalues and eigenvectors of the pencil \( S \). If we now put \( Z = -X - B \), the pencil factorises:

\[
S(\mu) = (1 - \mu Z)(1 - \mu X).
\]

The pencil \( S \) is said to be strongly damped if

\[
(Bx, x) > 2 \sqrt{\langle (x, x)(Cx, x) \rangle} \quad (x \neq 0).
\]

It is easy to see that in this case all the eigenvalues of the pencil are non-negative. Taking account of multiplicity, we shall denote by \((-\omega_n^*)\) those of them which are generated by the operator \( X \) and by \((-\omega_n^*)\) those generated by the operator \( Y \). It turns out that the pencil has no other eigenvalues (and there are no other roots of equation (44)). Each of the corresponding normalised systems of eigenvectors \( \{v_n^*\}, \{v_n^*\} \) is a Riesz basis in \( H \). In this sense the system of all eigenvectors of the pencil is doubly complete.

The phenomenon of multiple completeness of the system of root vectors of a polynomial operator pencil had already been discovered by M.V. Keldysh (1951). Since then the theory of polynomial pencils has developed into an
independent chapter of functional analysis and has found various applications in mathematical physics.

Remark. The investigation of quadratic equations in the algebra $\mathcal{L}(H)$ is, on the whole, a problem of transcendental difficulty. This is clear from the fact that for a continuously invertible operator $A$ the existence of a root of the equation $X^2 - AX = 0$ different from 0 and $A$ is equivalent to the existence of a non-trivial closed invariant subspace.

4.17. Banach Algebras. This important topic is closely connected with the spectral theory of operators. In fact, let $\mathfrak{H}$ be a complex commutative Banach algebra (in the sequel it will simply be called an algebra'). Let us consider the isometric embedding $T: \mathfrak{H} \to \mathcal{L}(\mathfrak{H})$ which associates with each $a \in \mathfrak{H}$ the operator $T_a x = ax$ ($x \in \mathfrak{H}$). Invertibility of an element $a$ is equivalent to invertibility of the operator $T_a$ in $\mathcal{L}(\mathfrak{H})$. Consequently, $\text{spec } a = \text{spec } T_a$ and so information about the spectra of bounded operators carries over to $\mathfrak{H}$. In particular, the spectrum of any element $a \in \mathfrak{H}$ is non-empty and compact, Gel'fand formula holds for the spectral radius $\rho(a)$ etc.

Example. Let $A$ be a bounded operator on a Banach space $E$ and let $[[A]]$ be the smallest closed subalgebra of $\mathcal{L}(E)$ which contains $A$, i.e. the uniform closure of the algebra of all polynomials in $A$. We denote the resolvent set and the spectrum of the element $A$ in the algebra $[[A]]$ by $\text{reg}_0 A$ and $\text{spec}_0 A$ respectively. Clearly, $\text{reg}_0 A \subset \text{reg } A$ and hence $\text{spec } A \subset \text{spec}_0 A$; moreover $\rho_0(A) = \rho(A)$ as a result of Gel'fand formula. It turns out that $\text{reg}_0 A$ coincides with the unbounded connected component $G_\infty$ of the set $\text{reg } A$; correspondingly $\text{spec}_0 A = \text{spec } A \cup G$, where $G$ is the union of the bounded connected components of the set $\text{reg } A$. In fact, if $|\lambda| > \rho(A)$ we see by expanding in a Laurent series at infinity that $R_\lambda \in [[A]]$. But then it follows from the Taylor expansion of the resolvent at the point $\lambda$ that $R_\mu \in [[A]]$ for $\mu$ in the maximal neighbourhood of the point $\lambda$ which does not intersect $\text{spec } A$. Connecting the point $\lambda$ to any point $\zeta \in G_\infty$ by means of a polygonal line, we deduce after a finite number of steps that $R_\zeta \in [[A]]$. Thus $G_\infty \subset \text{reg}_0 A$. Now let $H$ be a bounded connected component of the set $\text{reg } A$ and let $\mu \in H \cap \text{reg}_0 A$. Then for any $\varepsilon > 0$ there is a polynomial such that $\|R_\mu - p(A)\| < \varepsilon$. Any point $\lambda \in \partial H$ is contained in the approximate spectrum of the operator $A$. Let $(x_n)\varepsilon$ be a quasieigensequence: $Ax_n - \lambda x_n \to 0$, $\|x_n\| = 1$. Applying the operator $R_\mu - p(A)$ to $x_n$, we obtain in the limit: $|p(\lambda) - p(\mu)| < \varepsilon$. Thus the function $(\lambda - \mu)^{-1}$ can be uniformly approximated by polynomials on $\partial H$, which contradicts the existence of a singularity at the point $\mu \in H$.

It is now clear that the equality $\text{spec } A = \text{spec}_0 A$ holds if and only if $\text{spec } A$ does not disconnect the plane.

It is not difficult to determine that algebra defined by an operator $A$ for which the spectrum of the element $A$ coincides with that of the operator $A$. This is the smallest closed subalgebra $[[A]]$ of $\mathcal{L}(E)$ which contains $A$ and the family of operators $\{R_\lambda\}_{\lambda \in A}$, where $A$ is a set obtained by choosing one point from each bounded connected component of the set $\text{reg } A$ (the algebra $[[A]]$ does not depend on the choice of set $A$).
Returning to an arbitrary algebra $\mathcal{A}$, we note that for each closed ideal $J \subset \mathcal{A}$ the factor algebra $\mathcal{A}/J$ is a Banach algebra under its usual norm. But all maximal ideals are closed since the closure of an ideal $J \neq \mathcal{A}$ is an ideal $\overline{J} \neq \mathcal{A}$, as a consequence of the fact that the set of invertible elements is open. The factor algebra $E = \mathcal{A}/M$ for a maximal ideal $M$ is therefore a Banach field over $\mathbb{C}$. If $E \neq \mathbb{C}$ then for any $\xi \in E \setminus \mathbb{C}$ and for all $\lambda \in \mathbb{C}$ the non-zero elements $\xi - \lambda$ are invertible in $E$. But then $\text{spec } \xi = \emptyset$. With this we have established the Gelfand-Mazur theorem that a complex commutative Banach algebra is spectral.

The characters $\chi$ of an algebra $\mathcal{A}$ are continuous because of the closedness of their kernels, these being maximal ideals. Since $\chi(a) \in \text{spec } a$, we have $|\chi(a)| \leq \rho(a) \leq \|a\|$. Moreover $\chi(e) = 1$ and so $\|\chi\| = 1$.

Now let us consider the set $\mathcal{M} = \mathcal{M}(\mathcal{A})$ of maximal ideals of the algebra $\mathcal{A}$. It can be regarded as a subset of the unit sphere of the conjugate space $\mathcal{A}^*$, in which case it is closed since it is determined by the system of equations $\chi(a_1 a_2) = \chi(a_1) \chi(a_2) (a_1, a_2 \in \mathcal{A}, \chi(e) = 1$, which are defined by continuous functions of $\chi$. Consequently, $\mathcal{M}$ is compact.

**Example 1.** If $\mathcal{A} = C(S)$ ($S$ compact) then $\mathcal{M}(\mathcal{A}) = S$.

**Example 2.** If $\mathcal{A}$ is an algebra with a single topological generator $a$, i.e. $\mathcal{A} = [a]$, the uniform closure of the algebra of all polynomials in $a$, then $\mathcal{M}(\mathcal{A}) = \text{spec } a$.

We note that the problem of giving an explicit description of the maximal ideal space of a concrete algebra can be extremely difficult. An example of this is the famous problem which was solved by Carleson (1962); this is concerned with the algebra of functions which are analytic and bounded on the unit disk $\Delta = \{ \lambda : \lambda \in \mathbb{C}, |\lambda| < 1 \}$. The result is that $\Delta$ is dense in $\mathcal{M}$.

The canonical functional representation (Gelfand transformation) $a \mapsto \hat{a}$ ($\hat{a}(\chi) = \chi(a)$) is a morphism of $\mathcal{A} \to C(\mathcal{M}(\mathcal{A}))$ which is clearly contracting. For it to be an isometry (i.e. for the identity $\rho(a) = \|a\|$ to hold), it is necessary and sufficient that the identity $\|a^2\| = \|a\|^2$ should hold in $\mathcal{A}$.

The Gelfand image of the algebra $\mathcal{A}$ in $C(\mathcal{M})$ is a subalgebra (generally not closed) which separates the points of the space $\mathcal{M}$. It can fail to be dense in $C(\mathcal{M})$.

In the given context and also for many applications it is important to have an intrinsic characterisation of dense subalgebras of $C(S)$ ($S$ compact). The effective solution of this problem includes the Weierstrass approximation theorem (the subalgebra of polynomials is dense in $C[0, 1]$). It was obtained by Stone (1937).

**Stone-Weierstrass Theorem.** Suppose that a subalgebra $\mathcal{R} \subset C(S)$ is symmetric (i.e. along with each function $\varphi$ it contains its complex conjugate $\overline{\varphi}$) and separates the points of $S$. Then $\mathcal{R}$ is dense in $C(S)$.

For the proof (and also independently of this goal) it is useful (following de Branges, 1959) to deal first of all with the real case.

\[\text{In the category of complex commutative Banach algebras over } \mathbb{C} \text{ (i.e. a continuous algebra homomorphism) } \mathcal{M}(\mathcal{A}) \text{ is a contravariant functor into the category of compact spaces.}\]
Real Variant of the Stone-Weierstrass Theorem. Let $C_r(S)$ be the Banach algebra of real continuous functions on a compact set $S$ and let $\mathcal{R}_0$ be a subalgebra. If the functions in $\mathcal{R}_0$ separate the points of $S$, then $\mathcal{R}_0$ is dense in $C_r(S)$.

Proof. Suppose that the subalgebra $\mathcal{R}_0$ is not dense. Let us consider the set $N \neq 0$ of real measure on $S$ which annihilate $\mathcal{R}_0$ and satisfy the condition $||\nu|| \leq 1$. This is a convex, centrally symmetric, compact subset of the space of measures on $S$ under the $w^*$-topology. By the Krejn-Milman theorem $N$ has an extreme point $\sigma$. Clearly, $||\sigma|| = 1$ and by construction $\int_S \varphi \, d\sigma = 0$ ($\varphi \in \mathcal{R}_0$), in particular $\int_S d\sigma = 0$. Hence it follows that the support of the measure $\sigma$ contains at least two distinct points $s_1, s_2$. But there is a function $\psi$ in the subalgebra $\mathcal{R}_0$ which separates these points: $\psi(s_1) \neq \psi(s_2)$. Without loss of generality we can assume that $0 < \psi(s) < 1$ ($s \in S$). Since $\varphi \psi \in \mathcal{R}_0$ for all $\varphi \in \mathcal{R}_0$ we have $\int_S \varphi \psi \, d\sigma = 0$, i.e. the measure $d\tau = \psi \, d\sigma$ annihilates $\mathcal{R}_0$ and so the measure $d\rho = (1 - \psi) \, d\sigma$ also has this property. Putting $\xi = \tau/||\tau||$, $\rho = \rho/||\rho||$ we have $\sigma = p\xi + q\rho$ ($p, q > 0$, $p + q = 1$). Consequently $\xi = \alpha \sigma$ ($\alpha = \text{constant}$), i.e. $\psi|_{\text{supp} \sigma} = \text{constant}$.

For the proof of the complex Stone-Weierstrass theorem it is now sufficient to consider the set $\mathcal{R}_0 = \{\theta: \theta = \text{re} \varphi \ (\varphi \in \mathcal{R})\}$. If $\varphi \in \mathcal{R}$ we also have $\overline{\varphi} \in \mathcal{R}$ according to the condition and so $\text{re} \varphi \in \mathcal{R}_0$. Consequently, $\mathcal{R}_0 \subseteq \mathcal{R}$ and by the same token $\mathcal{R}_0$ is a subalgebra of the real algebra $C_r(S)$ and separates the points of the compact set $S$. According to the real Stone-Weierstrass theorem, $\mathcal{R}_0$ is dense in $C_r(S)$. But then $\mathcal{R}$ is dense in $C(S)$. \qed

A complex commutative Banach algebra $\mathcal{R}$ is said to be symmetric if its canonical image is symmetric.

Corollary. If the algebra $\mathcal{R}$ is symmetric, then its canonical image is dense in $C(\mathcal{M}(\mathcal{R}))$.

Remark. The Banach space $C_r(S)$ is also a vector lattice with respect to the usual ordering and moreover the two structures are compatible: if $||x|| \leq 1$ and $|y| \leq |x|$ ($|.|$ is the lattice modulus), then $||y|| \leq 1$. We can say that $C_r(S)$ is a Banach lattice. The following variant of the Stone-Weierstrass theorem holds: any sublattice of the vector lattice $C_r(S)$ which contains 1 and separates the points of the compact set $S$ is dense in $C_r(S)$.

Closed subalgebras of the algebras $C(S)$ ($S$ compact) are called uniform algebras. An example of such an algebra on the unit circle $T$ is the algebra $A(T)$ of functions which are the boundary values of analytic functions on the unit disk $\Delta$. It is called the disk algebra and it is maximal among the non-trivial closed subalgebras of $C(T)$ (Wermer, 1953). The maximal ideal space of the algebra $A(T)$ is naturally identified with the closed disk $\Delta = \Delta \cup T$. In general, if a uniform algebra $\mathcal{A}$ on $S$ separates points, then $S$ is homeomorphically embedded in the maximal ideal space $\mathcal{M}(\mathcal{A})$.

For any algebra $\mathcal{A}$ and any element $a \in \mathcal{A}$ we can construct, just as in the theory of operators, a functional calculus by putting
$\varphi(a) = -\frac{1}{2\pi i} \int_{\Gamma} \varphi(\lambda)(a - \lambda e)^{-1} d\lambda,$

where $\varphi$ is an analytic function on the spectrum, i.e. on some neighbourhood $G \supset \text{spec } a$, and the contour $\Gamma$ contains $\text{spec } a$ in its interior. Moreover $\varphi(\hat{a}(M)) = \varphi(\hat{a}(M)) (M \in \mathcal{M}(\mathfrak{M})), $ from which we obtain immediately the spectral mapping theorem:

$$\text{spec } \varphi(a) = \varphi(\text{spec } a).$$

**Example.** Let $W$ be the Wiener algebra, let $f \in W$ and let $\varphi$ be a function which is analytic on the set of values of the function $f$. Then the composition $\varphi \circ f$ belongs to $W$ (theorem of Levy; the case $\varphi(\zeta) = \zeta^{-1}$ is a theorem of Wiener).

An algebra $\mathfrak{A}$ is said to be regular if for any compact set $Q \subset \mathcal{M}(\mathfrak{A})$ and any maximal ideal $M_0 \notin Q$ there exists an element $a \in \mathfrak{A}$ such that $\hat{a}|_Q = 0$ and $\hat{a}(M_0) = 1$. It now follows from this that for any two disjoint compact subsets $Q_1, Q_2$ of $\mathcal{M}$ there exists an element $x \in \mathfrak{A}$ such that $\hat{x}|_{Q_1} = 0$ and $\hat{x}|_{Q_2} = 1$. If $\mathfrak{A}$ is symmetric, $x$ can be chosen so that $0 \leq \hat{x}(M) \leq 1$ for all $M$.

The theory of regular algebras was constructed by G.E. Shilov (1941). One of the fundamental results of this theory is the

**Local Theorem.** Suppose that the algebra $\mathfrak{A}$ is regular and that the function $\varphi$ belongs to $\mathfrak{A}$ locally on $\mathcal{M}$, i.e. for each point $M_0 \in \mathcal{M}$ there exists a neighbourhood $U$ and an element $a_0 \in \mathfrak{A}$ such that $\hat{a}_0(M) = \varphi(M) (M \in U)$. Then $\varphi \in \mathfrak{A}$. The point is that, because of the regularity of the algebra, for any finite open covering of the compact set $\mathcal{M}$ we can find a partition of unity as a sum of functions in $\mathfrak{A}$, each of which is concentrated on the corresponding element of the covering.

**Example.** If on a neighbourhood of each point a periodic function $\varphi$ can be expanded as an absolutely convergent trigonometric series (depending on the choice of point), then it has an expansion of this type on the whole period (Wiener, 1933).

An important consequence of the local theorem is a further generalisation of the Wiener-Levy theorem: suppose that the algebra $\mathfrak{A}$ is regular and that the function $\varphi$ on $\mathcal{M}$ is such that for each point $M_0 \in \mathcal{M}$ there exists a neighbourhood $U$, an element $a_0 \in \mathfrak{A}$ and a function $\varphi_0$ which is analytic on $\text{spec } a_0$ and such that $\varphi(M) = \varphi_0(\hat{a}_0(M));$ then $\varphi \in \mathfrak{A}$.

In certain cases this theorem allows us to apply multivalent analytic functions to elements of the algebra.

**Example.** If a function $a \in L^1(\mathbb{R})$ is such that its Fourier transform $A(\lambda)$ satisfies the conditions

$$1 - A(\lambda) \neq 0 \quad (\lambda \in \mathbb{R}), \quad \text{ind}(1 - A(\lambda)) = 0,$$

$^{119}$ An interesting variation on the topic of multivalent functions of an algebra element is the solution of algebraic equations with coefficients in the algebra. Non-trivial homomorphisms of the fundamental group $\pi_1(\mathfrak{M})$ into braid groups are global barriers to solvability (E.A. Gorin – V.Ya Lin, 1969).
then the function $\ln(1 - A(\lambda))$ is the Fourier transform of some function in $L^1(\mathbb{R})$. The role of this result in the theory of equations with a difference kernel is well-known (see Chapter 1, Section 4.5).

By way of explanation we state that for the Banach algebra $L^1(\mathbb{R})$ (with identity adjoined) the Gelfand transform and the Fourier transform coincide ($\tilde{I} = 1$). The maximal ideal space of this algebra (or of the Wiener algebra which is isometrically isomorphic to it) is naturally identified with the one-point compactification $\mathbb{R} = \mathbb{R} \cup \{\infty\}$.

In any algebra $\mathcal{A}$ an element $a$ is said to be logarithmic if it can be represented in the form $a = e^x$, where $x \in \mathbb{R}$. All such elements are clearly invertible but in general not all invertible elements are logarithmic.

**Example 1.** In the algebra $C[0, 1]$ all functions that are never zero are logarithmic, i.e. the functions which are invertible in the algebra.

**Example 2.** In the algebra $C(T)$ the logarithmic functions are precisely those non-vanishing functions $\varphi$ for which

$$\text{ind } \varphi = \frac{1}{2\pi} \Lambda_T(\arg \varphi) = 0.$$ 

We note that any invertible $\psi \in C(T)$ can be expressed in the form

$$\psi(\zeta) = \zeta^n \varphi(\zeta),$$

where $n = \text{ind } \psi$ and $\varphi$ is a logarithmic function.

The set $\exp \mathcal{A}$ of logarithmic elements of any algebra $\mathcal{A}$ is a subgroup of the group $\text{inv } \mathcal{A}$ of invertible elements. It turns out that the corresponding factor group depends only on the homotopy class of the maximal ideal space of the algebra.

**Arens-Royden Theorem.** The factor group $\text{inv } \mathcal{A}/\exp \mathcal{A}$ is isomorphic to the group $H^1(\mathcal{M}, \mathbb{Z})$ of one-dimensional Čech cohomologies of the maximal ideal space of the algebra $\mathcal{A}$.

In the proof of the Arens-Royden theorem (as also in many other questions in the theory of Banach algebras) we make use of analytic functions of $n$ algebra elements: $a_1, \ldots, a_n$. In the corresponding functional calculus$^{120}$ the role of the spectrum is played by the combined spectrum $\text{spec } a$ of the system $a = (a_1, \ldots, a_n)$, which is defined as the complement of the combined resolvent set. The latter consists of regular points $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$, for each of which there exists a system $b = (b_1, \ldots, b_n)$ of algebra elements such that

$$\sum_{k=1}^n (a_k - \lambda_k e)b_k = e.$$ 

It is easy to see that $\text{spec } a$ coincides with the image of the mapping $\tilde{a}: \mathcal{M} \to \mathbb{C}^n$

---

$^{120}$ It was developed in the works of G.E. Shilov (1953, 1960) and of Arens and Calderón (1955).
which is defined by the formula $\bar{a}(M) = (\bar{a}_1(M), \ldots, \bar{a}_n(M))$. Since this mapping is continuous, $\text{spec } a$ is compact. If it is polynomially convex, then, by the well-known Oka-Weil theorem, each function which is analytic on $\text{spec } a$ admits arbitrarily close uniform approximation by polynomials. This allows us to extend by continuity the intrinsic construction of the element $\varphi(a) = \varphi(a_1, a_2, \ldots, a_n)$ from the algebra of polynomials in $n$ variables to the algebra of functions which are analytic on the spectrum. A deeper analysis is required in the absence of polynomial convexity for the combined spectrum.

Returning to regular algebras, we note their significance for spectral theory.

**Example.** Let $\alpha = (\alpha_n)_{n \in \mathbb{Z}}$ be a real sequence such that

$$\alpha_n \geq 1, \quad \alpha_{n+m} \leq \alpha_n \alpha_m, \quad \sum_{n \in \mathbb{Z}} \ln \frac{\alpha_n}{n^2 + 1} < \infty.$$  

Then the Banach algebra $l^1_\infty$ (with convolution) of scalar sequences that are summable for the weight $\alpha$ is regular (G.E. Shilov, 1947). Its Gel'fand image, the Wiener algebra $\mathcal{W}_\alpha$ with weight $\alpha$, consists of those functions on the circle with Fourier coefficients in $l^1_\alpha$. The corresponding maximal ideal space is identified with the unit circle $T$.

If $V$ is a non-quasianalytic linear operator on a Banach space ($\text{spec } V \subset \mathbb{T}$), we can put $\alpha_n = \|V^n\|$. The functional calculus $q(V) = \sum_{n=-\infty}^{\infty} \alpha_n V^n$ arises on the algebra $\mathcal{W}_\alpha$. The spectrum of the operator turns out to be separated thanks to the fact that, for any compact set $Q \subset T$ and any neighbourhood $U \supseteq Q$, there exists a function $\varphi \in \mathcal{W}_\alpha$ such that $\varphi|_Q = 1$, $\varphi(T \setminus U) = 0$, $0 < \varphi < 1$ (clearly $\varphi$ is not only non-analytic but it also does not belong to any quasianalytic class).

We thus have convincing evidence that the methods of the theory of analytic functions are of paramount importance for the theory of Banach algebras (as indeed for spectral theory in general). Conversely, the theory of Banach algebras is a useful instrument for studying various classes of analytic functions, particularly in the setting of approximation theory. The concept of a boundary of the algebra plays an important role here. This term is applied to any subset $X$ of the maximal ideal space $\mathcal{M}$ such that for each function $a(M)$ ($a \in \mathcal{M}$) its maximum modulus is attained at some point $M_0 \in X$. The trivial boundary is $X = \mathcal{M}$ and, for example, in the case of the algebra $C(S)$ ($S$ compact) it is the only one. The boundaries for the algebra $A(T)$ are precisely those subsets of the disk $\Delta$ which contain the circle $T$. Since in any algebra an extension of a boundary preserves the boundary property, it is natural to look for a smallest boundary. G.E. Shilov (1940) proved that any algebra has a smallest closed boundary (subsequently called the Shilov boundary). Without the requirement of closedness a smallest boundary may not exist, however such a boundary does exist if the maximal ideal space is metrizable and in this case it consists of the peak points, i.e. those $M_0 \in \mathcal{M}$
for each of which there exists \( a_0 \in \mathcal{A} \) such that \( \tilde{a}_0(M_0) = 1 \) and \( |\tilde{a}_0(M)| < 1 \) \( (M \neq M_0) \). The results were obtained by Bishop (1959), who also gave the following application of them.

**Theorem.** A sufficient condition for the set \( R_0(S) \) of rational functions with poles outside of the compact set \( S \subset \mathbb{C} \) to be dense in \( C(S) \) is that (with respect to plane Lebesgue measure) almost every point of \( S \) be a peak point for some function in \( R(S) \) or \( R_0(S) \).

A.A. Gonchar (1963) showed that the following condition is sufficient for a point \( s_0 \) of a compact metric space to be a peak point of a uniform algebra \( \mathcal{A} \subset C(S) \): there exist constants \( c < 1, N \geq 1 \) such that for any neighbourhood \( U \) of \( s_0 \) there is a function \( \phi \in \mathcal{A} \) for which \( \phi(s_0) = 1, |\phi(s)| \leq c \) \( (s \in U) \) and \( \|\phi\| \leq N \). It follows from this that the set of peak points of the algebra \( \mathcal{A} \) is of type \( G_\alpha \).

Let \( \mathcal{A} \) be a complex Banach algebra (in general non-commutative). A mapping \( a \mapsto a^* \) of the algebra into itself is called an *involution* if
\[
\begin{align*}
1) & \quad (a + b)^* = a^* + b^*, \\
2) & \quad (ab)^* = b^*a^*, \\
3) & \quad (a^*)^* = a, \\
4) & \quad \|a^*\| = \|a\|. 
\end{align*}
\]
The element \( a^* \) is called the adjoint of \( a \). If \( a^* = a \), then \( a \) is said to be selfadjoint. If \( a^* \) commutes with \( a \), we say that \( a \) is normal. In particular, all unitary elements \( (a^*a = aa^* = e) \) are normal.

**Example.** The identity \( e \) is a selfadjoint unitary element.

If the element \( a \) is invertible, then \( a^* \) is also invertible and \( (a^*)^{-1} = (a^{-1})^* \).

Hence it follows that \( \text{reg } a^* = \overline{\text{reg } a} \) and \( \text{spec } a^* = \overline{\text{spec } a} \).

Any \( x \in \mathcal{A} \) can be uniquely represented in the form \( x = a + bi \), where \( a, b \) are selfadjoint: \( a \equiv \text{re } x = (x + x^*)/2, b \equiv \text{im } x = (x - x^*)/(2i) \). The selfadjoint elements of \( \mathcal{A} \) form a closed real subspace \( \text{re } \mathcal{A} \).

For any \( a \in \mathcal{A} \) the element \( a^*a \) is selfadjoint. The closure of the non-negative linear hull of the set \( \{x: x = a^*a \ (a \in \mathcal{A})\} \) is a wedge \( \mathcal{A}_+ \) whose elements, as usual, are said to be non-negative \( (x \geq 0) \). Using \( \mathcal{A}_+ \) we define in the canonical way the collection \( \mathcal{A}_+^\ast \) of non-negative linear functionals. Any \( f \in \mathcal{A}_+^\ast \) satisfies the inequality \( |f(a)| \leq f(e) \|a\| \) and is consequently continuous (\( \|f\| = f(e) \)). Moreover, it is real on \( \text{re } \mathcal{A} \) or equivalently: \( f(a^*) = \overline{f(a)} \ (a \in \mathcal{A}) \). Such functionals are said to be symmetric. Any functional \( g \in (\text{re } \mathcal{A})^\ast \) can be uniquely extended to a symmetric functional \( \tilde{g} \in \mathcal{A}^\ast \); also \( \|\tilde{g}\| = \|g\| \).

In the category of Banach algebras with involution the continuous homomorphisms \( h: \mathcal{A}_1 \rightarrow \mathcal{A}_2 \) which are compatible with the involutions, i.e. \( (ha)^* = ha^* \) (or, equivalently, \( h \) preserves selfadjointness of elements), are called morphisms (\( * \)-homomorphisms).

A subalgebra of an algebra with involution is said to be symmetric if the involution does not leave it. The class of symmetric ideals (one- or two-sided) is defined similarly. Under factorisation by a symmetric two-sided (one-sided) ideal the involution carries over naturally to the factor algebra (factor space).

---

\[\overline{\text{a}_0(M_0)} = 1\] Here the bar denotes complex conjugation.
A Banach algebra $\mathcal{A}$ with involution is called a $C^*$-algebra if

$$\|a^*a\| = \|a\|^2$$

for all $a \in \mathcal{A}$. We mention two examples of $C^*$-algebras.

**Example 1.** The algebra $C(S)$ of continuous functions on a compact set $S$ (the involution is complex conjugation).

**Example 2.** The algebra $\mathcal{L}(E)$ of bounded linear operators on a Hilbert space $E$ (the involution is the operator adjoint).

In fact, if $A \in \mathcal{L}(E)$ we have

$$\|A^*A\| = \sup_{\|x\|=1} (A^*Ax, x) = \sup_{\|x\|=1} \|Ax\|^2 = \|A\|^2.$$ 

Any closed symmetric subalgebra of a $C^*$-algebra clearly belongs to the same category. Any closed two-sided ideal of a $C^*$-algebra is symmetric and the corresponding factor algebra is a $C^*$-algebra.

**Gelfand-Najmark Theorem.** 1) Let $\mathcal{A}$ be a commutative $C^*$-algebra and $\mathcal{M}$ its maximal ideal space. Then the Gelfand representation is an isometric (\(\ast\)-)isomorphism.

2) Any $C^*$-algebra is isometrically (\(\ast\))-isomorphic to some operator algebra, i.e. a closed symmetric subalgebra of the algebra of operators $\mathcal{L}(E)$ for some Hilbert space $E$.

For the proof we note that for any $C^*$-algebra the spectra of unitary elements lie on the unit circle, since if $a^{-1} = a^*$ it follows from (47) that $\|a^{-1}\| = \|a\| = 1$, hence $\rho(a) \leq 1$ and $\rho(a^{-1}) \leq 1$; the spectra of selfadjoint elements are real, since if $a^* = a$ we have that $e^{ia}$ is unitary. Consequently, if $\mathcal{A}$ is commutative and $\chi$ is a character, then $\chi$ is real on $\mathcal{A}$ and hence it is symmetric. By the Stone-Weierstrass theorem $\mathcal{A}$ is dense in $C(\mathcal{M})$. It remains to show that $\rho(a) = \|a\|$ for all $a \in \mathcal{A}$. If $a = a^*$ this is obtained by iterations of the identity $\|a^2\|^2 = \|a\|^2$, while in the general case we have $\|a\|^2 = \|a^*a\| = \rho(a^*a) \leq \rho(a^*)\rho(a) = \rho^2(a)$. \(\square\)

The non-commutative situation is significantly more complicated and we will not enter into it.

Let us consider two applications of the preceding theorem.

**Example 1.** Let $S$ be a completely regular topological space. We will consider the commutative $C^*$-algebra $CB(S)$ on it. The canonical mapping of $S \to \mathcal{M}$ is a homeomorphism onto its image since $CB(S)$ separates points from closed subsets. We can therefore regard $S$ as a topological subspace of $\mathcal{M}$. But then $S$ is dense in $\mathcal{M}$ since, if $\varphi \in C(\mathcal{M})$ and $\varphi|_S = 0$, we have $\varphi = 0$ because of the isometric property of the Gelfand representation (which in this context becomes the non-commutative situation).

\[\text{\underline{122}}\text{We note in passing that the spectra of non-negative elements are non-negative.}\]
extension by continuity). Consequently $\mathcal{M}$ can be regarded as a compactification of $S$. If coincides with the Stone-Čech compactification. All possible compactifications are in one-to-one correspondence with the closed symmetric subalgebras $\mathcal{A} \subset CB(S)$ which separate points. However the topology of the compactification $\mathcal{M}(\mathcal{A})$ induces on $S$ a topology which is coarser than its initial topology and may not coincide with it. For example, the Bohr compactification of the space $\mathbb{R}$, which is defined by the algebra $AP(\mathbb{R})$ of almost periodic functions, has this property.

**Example 2.** Let $A$ be a bounded normal operator on a Hilbert space. Then $A$ and $A^*$ generate a commutative $C^*$-algebra (this is the smallest closed subalgebra of $\mathcal{B}(E)$ which contains $A$ and $A^*$). It can be shown that the spectrum of the operator $A$ coincides with its spectrum as an element of this algebra $\mathcal{A}$ and can be identified with its maximal ideal space $\mathcal{M}$. The Gel'fand representation of $\mathcal{M} \to C(\mathcal{M})$ is an isometric ($*$)-isomorphism. The inverse mapping can be regarded as a functional calculus. It can be extended (by continuity in the topology of pointwise convergence) to the algebra of bounded Baire functions $\varphi(\lambda)$ ($\lambda \in \text{spec } A$). The orthoprojections forming an orthogonal resolution of the identity for the operator $A$ correspond to the characteristic functions of the Baire subsets of the spectrum. Thus in this way we obtain the spectral theorem for a bounded normal operator.

Banach algebras with involution play a significant role in the theory of representations and harmonic analysis. If $G$ is a locally compact group and $\mu$ is a right Haar measure, the convolution operation

$$(\varphi * \psi)(g) = \int_G \varphi(\gamma)\psi(g\gamma^{-1}) \, d\mu(\gamma) \quad (g \in G)$$

turns $L^1(G, \mu)$ into a Banach algebra (without identity if $G$ is not discrete; we then have to adjoin an identity formally. This group algebra is commutative if and only if the group $G$ is commutative and in this case its characters can be identified with the unitary characters of the group $G^*$ by means of the Fourier transform

$$\hat{\varphi}(\chi) = \int_G \varphi(g)\chi(g) \, d\mu(g) \quad (\chi \in G^*)$$

(if $G$ is not discrete, the group algebra has another character – the coefficient of the identity). The maximal ideal space of the group algebra coincides with $G^*$ if $G^*$ is compact ($\iff G$ is discrete), while in the general case it is the one-point compactification of $G^*$. The Gel'fand representation for the group algebra $L^1(G, \mu)$ coincides with the Fourier transform, i.e. for $G = \mathbb{R}$ we have the classical Fourier transform and for $G = \mathbb{T}$ it associates with each function its sequence of Fourier coefficients. We have correspondingly $\mathbb{T}^* = \mathbb{Z}$ and $\mathbb{R}^* = \mathbb{R}$. The group

\[123\text{These form the dual group } G^* \text{ (pointwise multiplication). The usual topology on } G^* \text{ is the compact-open topology, under which } G^* \text{ is locally compact.}\]
algebra of a locally compact abelian group is semisimple, which is equivalent to the uniqueness theorem for the Fourier transform.

An involution can be introduced on a group algebra by means of the formula \( \varphi^*(g) = \varphi(g^{-1}) \). Although this is not a C*-algebra, the presence of an involution on it nevertheless turns out to be extremely useful, for example, in such a central question as the generalisation of Plancherel's theorem\(^{24}\) (A. Weil, 1940; M.G. Krejn, 1941). In this way D.A. Rajkov (1941) obtained a new proof of the Pontryagin duality principle, according to which for any locally compact abelian group the canonical mapping of \( G \to G^{**} \) is a topological isomorphism\(^{125}\).

The principle of duality contains, in particular, the assertion that the unitary characters of a locally compact abelian group separate its points (in this sense there are sufficiently many of them). This result extends to the non-commutative situation if we replace characters by irreducible unitary representations (Gelfand-Rajkov theorem). Moreover, any unitary representation of a locally compact group can be expressed as the direct integral of irreducible unitary representations (M.A. Najmark, 1956). This theorem bears the same relationship to the previous one as Choquet's theorem does to the Krejn-Milman theorem and the technique of extreme points is actually used in the corresponding proofs. For a compact group the expression reduces to a discrete orthogonal sum (with a suitable modification this is also true for representations in a Banach space). For the regular representation of a compact group the decomposition into an orthogonal sum of irreducibles is a basic result from the Peter-Weyl theory. We note that all irreducible representations of a compact group (even in a locally convex space) are finite-dimensional and can therefore be taken to be unitary.

Compact groups (even non-abelian) have their own particular principle of duality (Tannaka, 1939; M.G. Krejn, 1949); the object which is dual to the group is the so-called block-algebra.

The study of any Banach algebra \( \mathcal{A} \) with involution reduces to the case of a C*-algebra in the sense that there exists a C*-algebra \( \hat{\mathcal{A}} \), unique up to (**) isomorphism, which is associated with \( \mathcal{A} \) by means of a morphism \( \rho: \mathcal{A} \to \hat{\mathcal{A}} \) in such a way that, for any morphism \( h: \mathcal{A} \to \mathcal{B} \) (\( \mathcal{B} \) a C*-algebra), we have the commutative diagram

\[ \begin{array}{ccc} \mathcal{A} & \xrightarrow{\rho} & \hat{\mathcal{A}} \\ \downarrow h & & \downarrow \hat{h} \\ \mathcal{B} & \xrightarrow{\rho^*} & \hat{\mathcal{B}} \end{array} \]

\(^{125}\) A suitably normalised' Haar measure \( \mu^* \) on \( G^* \) appears in the inversion formula

\[ \varphi(g) = \int_{G^*} \phi(\chi) \chi(g) \, d\mu^*(\chi) \quad (\hat{\psi} \in \hat{L}^1(G^*, \mu^*)) \]

and correspondingly in the generalised Plancherel theorem. In this context it is appropriate to note also a generalisation of Bochner's theorem on the representation of positive functions:

\[ \varphi(g) = \int_{G^*} \chi(g) \, d\sigma(\varphi) \]

where \( \sigma \) is a finite measure (\( \sigma \geq 0 \)).

\(^{24}\) The original proof (L.S. Pontryagin, 1934) was based on structure theory. The requirement contained in it of a countable base was subsequently removed (van Kampen, 1935).
where $\hat{h}$ is some morphism. The $C^*$-algebra $\hat{\mathfrak{A}}$ is said to be universally enveloping for $\mathfrak{A}$. This is an extremely useful construction. For example, it works effectively in the proof of the Gel'fand-Rajkov theorem (where $\mathfrak{A}$ is a group algebra).

In the 1950s, Yu. M. Berezanskij and S.G. Krejn extended the scope of harmonic analysis, defining the generalised convolution on $L^1(S, m)$ for any locally compact space $S$ with measure $m$:

$$
(\varphi \ast \psi)(t) = \int_S \varphi(s) d_s \left( \int_S \psi(r) d_r C(s, r; t) \right) \quad (t \in S),
$$

where for fixed $t$ $C(s, r; t)$ is a Borel measure with respect to $s$ and $r$ separately and for fixed subsets in place of $s$ and $r$ it is a continuous function of $t$ with compact support; it is required in addition that

$$
\int_S C(A, B; t) \, dm(t) \leq m(A)m(B).
$$

We then have $\|\varphi \ast \psi\| \leq \|\varphi\| \|\psi\|$ and so a commutative Banach algebra with structure measure\footnote{This is the continuous analogue of the structure constants of finite-dimensional algebra. It is assumed that relationships hold which guarantee associativity and commutativity. An identity may be lacking but it can be adjointed formally in the usual way.} $C(A, B; t)$ arises on $L^1(S, m)$. It is assumed further that an involutory homomorphism $s \mapsto s^*$ acts on $S$, preserving the measure $m$ and acting on the structure measure so that an involution is defined on the Banach algebra under consideration by the formula $\varphi^*(t) = \overline{\varphi}(t^*)$. From these stand-points, for example, ordinary (group) and generalised (hypergroup) shifts can be given a unified treatment.

§ 5. Function Spaces

5.1. Introductory Examples. The development of functional analysis is connected not only with the working out of its abstract intrinsic structures but also with the expansion of the stock of concrete function spaces. This has both an applied and a conceptual meaning. The revolutionary introduction into mathematical usage of the language and techniques of generalised functions can serve as a striking example (S.L. Sobolev, 1936; L. Schwartz, 1950–51). The general theory of interpolation of linear operators, which came into being right at the
beginning of the 1960s (Gagliardo, Lions, Calderón, S.G. Krejn), developed on
the one hand from concrete material relating to spaces of type $L^p$, while on the
other hand it opened up a broad field of activity in new function spaces (for
example, in symmetric spaces of measurable functions). Interpolation methods
are intrinsically linked with embedding theorems of S.L. Sobolev, which are
concerned with spaces of differentiable functions of several variables. In connection
with multivariate approximation theory important advances were made in
the theory of embeddings during the 1950s thanks to work of S.M. Nikolskij
and O.V. Besov, in which new classes of differentiable functions were introduced,
leading, in particular, to valuable converses of embedding theorems. At the
present time we can say that function spaces form an independent topic, in the
investigation of which abstract and concrete analytical methods are intertwined.

The successful solution of many problems in the theory of partial differential
equations depends on the right choice of function space (or family of spaces)
appropriate for the given problem.

Example 1. The Cauchy problem for the heat equation

$$\frac{\partial u(x, t)}{\partial t} - \frac{\partial^2 u(x, t)}{\partial x^2} \quad (t \geq 0, \ -\infty < x < \infty), \quad u(x, 0) = \varphi(x).$$

is initially set in the class of all sufficiently smooth functions but in this situation
the solution is not unique. In fact in the non-quasianalytic class $C(n^{1[0, 1]})$
$(0 < \varepsilon < 1)$ on the interval $[0, 1]$ there exists a function $\psi(t)$ which together with
all its derivatives vanishes at the end points. Let us put $\psi(t) = 0$ for $t > 1$. Then

$$u(x, t) = \sum_{n=0}^{\infty} \frac{\psi^{(2n)}(t)}{(2n)!} x^{2n} \quad (t \geq 0, \ -\infty < x < \infty)$$

is a solution of the heat equation and $u|_{t=0} = 0$. The reason for this phenomenon
lies in the possibility of rapid growth of $u(x, t)$ as $|x| \to \infty$. Täkläind (1937) showed
that if $h(x)$ is a non-decreasing odd function then the bound

$$|u(x, t)| \leq M e^{\varepsilon h(x)}$$

defines a uniqueness class if and only if

$$\int_{1}^{\infty} dx \frac{x}{h(x)} = \infty. \quad (48)$$

This result is closely connected with a criterion for quasianalyticity. In particular,
condition (48) is satisfied by the function $h(x) = x$, which defines the well-known uniqueness class $|u(x, t)| \leq M e^{x^2}$ identified by A.N. Tikhonov (1935).

Example 2. In the Dirichlet problem for Laplace equation

$$\Delta u = 0, \quad u|_{\Gamma} = \varphi$$

in a domain $G \subset \mathbb{R}^n$ with (sufficiently well-behaved) boundary $\Gamma$ a key role is
played by the following quadratic functional (known as the Dirichlet integral):
\[ D[u] = \int_G |\nabla u|^2 \, dx \]

\((dx \text{ is Lebesgue measure})\). The Hilbert space of functions \(W^1_2(G)\) with norm

\[ \|u\| = \sqrt{D[u] + \|u\|_{W^2_2(G)}} \]

(introduced by S.L. Sobolev) is intrinsically connected with this problem. This norm requires the existence of partial derivatives if only almost everywhere, however this is not enough. An adequate description of the space is obtained by the introduction of \textit{generalised derivatives} (see below).

Since \(\Delta u = 0\) is an Euler-Lagrange equation for the functional \(D[u]\), it is appropriate to change the formulation of the Dirichlet problem by requiring that 

\[ D[u] = \min, \quad u|_\Gamma = \varphi, \quad u \in W^1_2(G). \]

The question of second order smoothness is thereby separated from the questions of existence and uniqueness (after which it can be solved by methods specially adapted for it). However the variational formulation also contains a subtle aspect which concerns the boundary condition: an admissible function \(\varphi\) must be the limit of functions in \(W^1_2(G)\) in a certain sense (requiring more precise definition) and it is desirable to have an intrinsic characterisation of the class of these functions. The answer to this question requires the introduction of the new function space \(W^{1/2}_1(\Gamma)\) of functions which are smooth of order \(\frac{1}{2}\) in a certain integral sense. It turns out that the restriction homomorphism \(u \mapsto u|_\Gamma\) acting from \(W^1_2(G)\) into \(W^{1/2}_1(\Gamma)\) is continuous and surjective\(^1\) (Aronszajn, 1955; V.M. Babich – L.N. Slobodetskij, 1956).

\[ 5.2. \text{Generalised Functions.} \]

The simplest generalised function (although not a function in the usual sense) is the \(\delta\)-function which is in fact a continuous linear functional on the \(\text{basic}\) space of continuous functions \(C(\mathbb{R})\) (for the time being we restrict ourselves for simplicity to the one-dimensional case):

\[ \delta[f] = f(0), \]

and formally

\[ \int_{-\infty}^{\infty} \delta(x) f(x) \, dx = f(0). \]

By formally integrating by parts we can extract from this equality an heuristic definition of the derivative \(\delta'\):

\[ \int_{-\infty}^{\infty} \delta'(x) f(x) \, dx = -\int_{-\infty}^{\infty} \delta(x) f'(x) \, dx = -f'(0). \]

Thus we can define a linear functional \(\delta'[f] = -f'(0)\), however \(C^1(\mathbb{R})\) will be the natural basic space for this (and \(\delta'\) turns out to be continuous). For the simultaneous definition of all derivatives \(\delta^{(n)}[f] = (-1)^n f^{(n)}(0)\) \((n \geq 1)\) we have to take

\(^{1}\) The boundary function \(\varphi = u|_\Gamma\) is defined almost everywhere and is the limit of \(u\) in a certain integral sense.
C^\infty(\mathbb{R}) as the basic space. On the other hand, however, the δ-function is an elementary measure, namely the unit measure concentrated at zero.\textsuperscript{2} If we wish to regard all measures as generalised functions, then, because a measure can grow arbitrarily at infinity, we have to take for the basic space C_0(\mathbb{R}), the space of continuous functions with compact support. C^\infty_0(\mathbb{R}), the space of infinitely differentiable functions with compact support, is the universal receptacle of all the formulated requirements. It is clear from what has been said that a generalised function has to be defined not as an individual object but as an element of some class.

Let G be an open set in \mathbb{R}^n and E some linear topological space of functions f: G \to \mathbb{C} which we will call the basic functions. The continuous linear functionals on E, i.e. the elements of the conjugate space E^* (with its w*-topology), are called generalised functions over E.

The first and most important example is E = C^\infty(G) = \mathfrak{D}(G) (to say that f: G \to \mathbb{C} has compact support means that there is a compact set Q_f \subset G outside of which f(x) = 0; the smallest such compact set, supp f, is the support of the function f). We introduce the locally convex (non-metrizable) Dieudonné-Schwartz topology on \mathfrak{D}(G) by means of the family of seminorms

\[ P_\rho(f) = \sum_{k_1, \ldots, k_n=0}^{\infty} \sup_{x \in G} \rho_{k_1, \ldots, k_n} \left| \frac{\partial^{k_1+\cdots+k_n} f}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}} \right| , \]

where \rho = (\rho_{k_1, \ldots, k_n}) is an arbitrary n-multisequence of continuous functions on G which is locally finite in the sense that for each compact set Q \subset G the set of multi-indices \( k = (k_1, \ldots, k_n) \) for which \( \rho_{k_1, \ldots, k_n} \neq 0 \) is finite.

With \mathfrak{D}(G) as the choice of basic space the generalised functions are called distributions. Thus \( \lambda \) is a distribution if it is a linear functional on \mathfrak{D}(G) which is subordinate to a finite family of seminorms \( p_{\rho^{(1)}}, \ldots, p_{\rho^{(n)}} \). An effective criterion for continuity of a functional \( \lambda \in \mathfrak{D}(G)^* \) is that for each compact set Q \subset G there exists an integer \( m = m(Q) \) (we shall assume it to be as small as possible) and \( C = C(Q) > 0 \) such that

\[ |\lambda[f]| \leq C \sum_{|k| \leq m} \sup_{Q} \left| \frac{\partial^k f}{\partial x^k} \right| \]

for all \( f \in \mathfrak{D}(G) \) with supp \( f \subset Q \). If \( \sup_{Q} m(Q) = r < \infty \), we say that \( \lambda \) is a distribution of finite order \( r \). A distribution of order zero is a complex measure. We can also regard a locally summable function \( \theta \) as a distribution by identifying it with the measure \( \int \theta \, dx \). Such distributions are said to be regular and form a

\[ \text{regular} \]

is convenient to interpret the δ-function physically as a unit charge concentrated at zero. Then \( \delta' \) is a dipole concentrated at zero with unit momentum. However in quantum mechanics the concept of the δ-function is essentially a functional.

The use of multi-indices in work with functions of n variables is very convenient since it leads to abbreviations of the form: \( x^k = x_1^{k_1} \cdots x_n^{k_n} \), \( \frac{\partial^k f}{\partial x^k} = \frac{\partial^{k_1+\cdots+k_n} f}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}} \). It is also convenient to put \( |k| = k_1 + \cdots + k_n \).
dense subspace $L^1_{\text{loc}}(G) \subset \mathcal{D}(G)^\ast$. In general it is far from being the case that we can speak of the values of a distribution $\lambda$ at points $x \in G$ (although the conventional notation $\lambda(x)$ is widely used because of its formal convenience\(^4\)). If an open set $H \subset G$ is given and $\lambda[f] = 0$ for all $f \in \mathcal{D}(H)$, we set\(^5\) $\lambda|_H = 0$. The complement in $G$ of the union of all such $H$ is called the support of the distribution $\lambda$ and is denoted by $\text{supp} \lambda$. The distributions with compact support form a space which is the conjugate of $C^\infty(G)$. The topology on $C^\infty(G)$ is defined by the collection of seminorms appearing in (49), i.e. it is the topology of uniform convergence along with all derivatives on each compact set. Clearly, $C^\infty(G)$ is a locally convex Fréchet space.

Remark 1. Without going into the general definitions we note that $\mathcal{D}(G)$ is the inductive limit of the family $C^\infty(X)$, where $X$ runs through all closed balls contained in $G$. The topology on $\mathcal{D}(G)$ is stronger than that induced by $C^\infty(G)$.

Remark 2. If $\lambda, \mu \in \mathcal{D}(G)^\ast$, $H \subset G$ is an open set and $(\lambda - \mu)|_H = 0$, then we say that the distributions $\lambda, \mu$ coincide on $H$. If $\theta \in L^1_{\text{loc}}(H)$ and $\theta|_H = \theta$, then the values $\lambda(x)$ are defined at all Lebesgue points of the function $\theta$.

Selecting the cone of (pointwise) non-negative functions in $\mathcal{D}(G)$ we obtain the cone of non-negative distributions in $\mathcal{D}(G)^\ast$. It turns out that it coincides with the cone of measures.

Multiplication by a function $\varphi \in C^\infty(G)$ is a continuous linear operator\(^6\) on $\mathcal{D}(G)$. Its conjugate operator on $\mathcal{D}(G)^\ast$ is also called an operator of multiplication by $\varphi$: $(\varphi \lambda)[f] = \lambda[\varphi f]$. For example, $\varphi \delta = \varphi(0)\delta$. This operation does not allow extension to multiplication in $\mathcal{D}(G)$, which would convert $\mathcal{D}(G)^\ast$ into a commutative algebra.\(^7\) Thus $\mathcal{D}(G)^\ast$ is only a module over the algebra $C^\infty(G)$.

The differentiations $D_j = \partial/\partial x_j (1 \leq j \leq n)$ are also continuous linear operators on $\mathcal{D}(G)$ and consequently the conjugate operators $D_j^\ast$ act on $\mathcal{D}(G)^\ast$, however, as already explained by the example of the $\delta$-function, it follows that we should put $\partial \lambda/\partial x_j = -D_j^\ast \lambda$ for any distribution $\lambda$.

\(^4\)For example, in this form it is convenient to make the change of variables $x = Fy$ ($F: G \to G$ is a diffeomorphism):

$$\int_G \lambda(Fy)f(y) \, dy = \int_G \lambda(x)f(F^{-1}x) \frac{dx}{\det F'(F^{-1}x)}.$$  

Consequently, it is necessary to take as definition

$$(\lambda \circ F)[f] = \lambda \left[ \frac{f \circ F^{-1}}{\det F'(F^{-1})} \right].$$

\(^5\)The natural embedding of $\mathcal{D}(H)$ into $\mathcal{D}(G)$ (extension by zero) is a continuous homomorphism whose conjugate is the homomorphism which restricts a functional to $\mathcal{D}(H)$.

\(^6\)Continuous operators of multiplication on function spaces by a function (possibly not belonging to the given space) are called multiplicators.

\(^7\)This circumstance can be regarded as a formal cause of the appearance of divergence in quantum field theory and the removal of divergence (i.e. normalisation of the theory) as the regularisation of multiplication of generalised functions (N.N. Bogolyubov – O.S. Parasyuk, 1955).
All linear differential operators with coefficients in $C^\infty(G)$ are now defined on $\mathcal{D}(G)^*$:

$$L[\lambda] = \sum_{|k| \leq m} a_k(x) \frac{\partial^{k\lambda}}{\partial x^k}.$$ 

They are all continuous under the $w^*$-topology, which is taken as the standard topology on $\mathcal{D}(G)^*$.

As an example to illustrate the flexibility and convenience of the language of distributions, let us mention the general definition of one of the principal concepts of the theory of partial differential equations. For any linear differential operator $L$ with coefficients in $C^\infty(G)$ a distribution $\Gamma_y$, which depends on the parameter $y \in G$ and satisfies the equation

$$L[\Gamma_y] = \delta_y,$$

is called a fundamental solution. For example, the Laplace operator in $\mathbb{R}^n$ has as a fundamental solution when $n > 2$ the function

$$\Gamma_y(x) \equiv \Gamma(x, y) = -\frac{1}{(n-2)\sigma_n|x-y|^{n-2}},$$

where $\sigma_n$ is the area of the unit sphere of $\mathbb{R}^n$ with the canonical Euclidean norm

$$|x| = \sqrt{\sum_{j=1}^n x_j^2}.$$ 

A fundamental solution $\Gamma_y$ exists for any operator $L$ with constant coefficients ( Ehrenpreis, 1954; Malgrange, 1955) and clearly it has the form $10$ $\Gamma_y(x) = \Gamma(x - y)$ ($x, y \in \mathbb{R}^n$). An immediate application of this fundamental solution is the construction of solutions to the equation $11$ $L[\lambda] = \mu$ by means of the formula $\lambda = \Gamma \ast \mu$. Here we are using convolution of distributions. For its introduction it is necessary to consider first the convolution of the distribution $\mu$ with an arbitrary function $f \in \mathcal{D}(\mathbb{R}^n)$, i.e. the function

$$(\mu \ast f)(x) = \mu[f(x - \cdot)].$$

This function belongs to $C^\infty(\mathbb{R}^n)$ but if the distribution $\mu$ has compact support

---

$8$ $\delta_y$ is the unit measure concentrated at the point $y$. In $\mathbb{R}^n$ we can write $\delta_y(x) = \delta(x - y)$. This notation carries over to any domain $G$.

$9$ If $n = 2$ we have $\Gamma(x, y) = \frac{1}{2\pi} \ln|x|$. If $n = 3$ we can interpret $\Gamma(x, y)$ physically as the potential of a point charge.

$10$ In a more complete formulation of the theorem it is asserted that $\Gamma$ is a distribution of finite order. We note that in theorems of this type not only the formulations are important but also the constructions which allow us to establish further useful properties of the solutions. The generalised Fourier transform is usually applied in the construction of fundamental solutions of equations with constant coefficients (see below).

$11$ In the case where $L$ is the Laplace operator this is Poisson's equation describing the potential of a charge with density $\mu$. 

---
then \( \mu \ast f \in \mathcal{D}(\mathbb{R}^n) \) and, by definition, we can put \((\Gamma \ast \mu)[f] = \Gamma \ast (\mu \ast f)\). Thus the equation \( L[\lambda] = \mu \) is successfully solved (in the class of all distributions) for any distribution \( \mu \) with compact support. In particular, \( \mu \) can be a summable function with compact support but the solution \( \lambda \) nevertheless turns out to be only a distribution in general (generalised solution). However, if the operator \( L \) is elliptic\(^{12}\) (and only in this case), all generalised solutions will be real analytic functions whenever the right hand side of the equation is a real analytic function (I.G. Petrovskij, 1937; Hörmander, 1955). An analogous result relating to infinite differentiability holds for the so-called hypoelliptic operators\(^{13}\) (and only in this case) (Hörmander, 1955). Generalised solutions are entirely natural from the physical point of view since they satisfy the equation in an integral sense: the equality \( L[\lambda] = \mu \) means that \( \lambda[L_0 f] = \mu[f] \) for all \( f \in \mathcal{D}(G) \), i.e. in the usual notation

\[
\int_G \lambda(x)(L_0 f)(x) \, dx = \int_G \mu(x)f(x) \, dx \quad (L_0^* = L).
\]

This form of the equation can have even greater physical meaning than the traditional one.

The conjugate operator principle' has already been used repeatedly to transfer the standard operations of analysis to generalised functions: if a continuous linear operator \( A \) is given on the basic space, then \( A^* \) is its transfer to the space of generalised functions. In a more general situation we can be given a continuous homomorphism \( h: E \rightarrow E_1 \) of the basic space into another function space. Since \( h \ast \mu \in E^* (\mu \in E_1^*) \), it follows that \( h \ast \mu \) is a generalised function (on \( E \)). We can regard this as the transfer of the homomorphism \( h \) to the generalised functions over the basic space \( E_1 \). The most important example of this type is the generalised Fourier transform.

Suppose that the basic space \( E \) consists of functions on \( \mathbb{R}^n \) and \( E \subset L^1(\mathbb{R}^n) \). Then the ordinary Fourier transform\(^{14}\)

\[
f(\xi) = \int_{\mathbb{R}^n} f(x) e^{-i(\xi, x)} \, dx
\]

is defined on \( E \). It maps \( E \) bijectively to a certain function space \( \hat{E} \) (contained in the space of continuous functions which tend to zero at infinity). Let us provide \( \hat{E} \) with the topology carried over from \( E \). As a result the Fourier transform can be transferred to the generalised functions on the basic space \( E \). For example, the Fourier transform of any distribution on \( \mathbb{R}^n \) is a generalised function over the basic space of entire (after analytic continuation with respect to \( x \)) functions of

\(^{12}\) An operator \( L \) is said to be elliptic if its principal part \( \sum_{|\alpha|=n} a_\alpha \frac{\partial^\alpha}{\partial x^\alpha} \) is defined by a homogeneous polynomial \( P(\xi_1, \ldots, \xi_n) \) (symbol) with no real roots \( \xi \neq 0 \).

\(^{13}\) All elliptic operators are hypoelliptic. Infinite differentiability of solutions of elliptic equations with infinitely differentiable right hand sides was established by L. Schwartz (1951).

\(^{14}\) \((\xi, x)\) is the canonical scalar product on \( \mathbb{R}^n \).
exponential type that decrease more rapidly than any power of $|x|$. In particular we can now speak of the Fourier transform of any locally summable function $\theta(x)$ without imposing any restrictions as $|x| \to \infty$.

It is clear that the generalised Fourier transform has the usual 'operational' properties: differentiation $D_j$ passes over to multiplication by $i\xi^j$, therefore any linear differential operator with constant coefficients passes over to the operator of multiplication by the corresponding polynomial; convolution becomes a product and so on; hence differential, convolution and some other equations are as usual transformed into algebraic equations. For example, the equation $L[\lambda] = \mu$ takes the form $P(\xi)\lambda = \mu$ and the problem reduces to determining the possibility of dividing a generalised function by a polynomial (division problem).

The general approach to the Fourier transform was developed gradually (Bochner, 1932; L. Schwartz, 1951) and took its final form, which has been described above, in the work of I.M. Gel'fand and G.E. Shilov (1953), where they obtained with the help of the Fourier transform uniqueness classes for the Cauchy problem for systems of differential equations of the form

$$\frac{\partial u_j}{\partial t} = \sum_{k=1}^m L_{jk}(t)[u_k] \quad (1 \leq j \leq m),$$

the $L_{jk}(t)$ being linear differential operators in $\mathbb{R}^n$ with coefficients which depend continuously on $t$ (but are constant with respect to $x$). If $p$ is the least order of the operators $L_{jk}$, then the uniqueness class is defined by the bound

$$\max_j |u_j(x, t)| \leq Me^{\sigma|x|^q},$$

where $q = p/(p - 1)$ and $\sigma, M$ are positive constants. In this connection the consideration of generalised functions over suitable spaces of entire functions turned out to be essential, since it allowed the possibility of establishing uniqueness by using the Phragmen-Lindelöf principle. The application of this sort of analytic technique requires an intrinsic characterisation of the space $E^*$ for those spaces $E$ being used as basic spaces.

We stress the lack of sensitivity of the method described to the type of system (for parabolic (according to Petrovskij) equations the uniqueness class was established by O.A. Ladyzhenskaya (1950) and for parabolic systems by S.D. Ehjdéman (1953); in this situation G.H. Zolotarev (1958) obtained a criterion of Taklind type). The general result is well illustrated by the following example: for the equation

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2},$$

$^{15}$In fact we can take $q = \frac{p_0}{p_0 - 1}$, where $p_0 \leq p$ is the so-called reduced order. There is an effective formula for its determination, from which we can see that $p_0$ is a rational number with denominator $\leq m$; in particular, $p_0 = p$ if $m = 1$ (V.M. Borok, 1957).

$^{16}$This is a certain (not very great) cost on account of the restrictions at infinity.
with complex $a$ the uniqueness class in the Cauchy problem is defined by the bound

$$|u(x, t)| \leq Me^{\alpha x^2}. \quad (52)$$

For $a > 0$ this is the heat equation, for $a < 0$ it represents 'reverse' heat flow and for $a = i$ we have the simplest case of a non-stationary Schrödinger equation.

We note that for $a < 0$ the Cauchy problem for equation (51) is incorrect in the sense that we do not have continuous dependence of the solution (in the uniform metric on an arbitrarily small time interval) on the initial function.\footnote{The physical significance of this fact is that the process of heat conduction is not reversible in time.}

The technique of generalised Fourier transform allows us to identify correctness classes in the Cauchy problem for the system (50) (G.E. Shilov, 1955). Certainly they are, in general, no longer the uniqueness classes, although sometimes the two coincide. For example, the Cauchy problem for the heat equation in the class (52) is correct.

Now let the basic space $E$ be an extension of the space $\mathcal{D} = \mathcal{D}(\mathbb{R}^n)$ with a weakened topology and also let $\mathcal{D}$ be dense in $E$ for the weak topology. Then all continuous linear functionals on $E$ remain such when restricted to $\mathcal{D}$ and are uniquely defined by their restrictions. Thus generalised functions on $E$ can be regarded in the given case as distributions. The most remarkable example of this type is the Schwartz space $\mathcal{S}$ of functions of class $C^\infty(\mathbb{R}^n)$ which decrease together with all their derivatives more rapidly than any power $|x|^{-v}$ ($v = 0, 1, 2, \ldots$). This is a locally convex Fréchet space defined by the family of seminorms

$$p_{k,v}(f) = \sup_{x \in \mathbb{R}^n} (1 + |x|^v) \left| \frac{\partial^k f}{\partial x^k} \right|.$$ 

All linear differential operators with polynomial coefficients act continuously on $\mathcal{S}$. The Fourier transform is a topological automorphism of the space $\mathcal{S}$.

Generalised functions on $\mathcal{S}$ are called tempered distributions. All linear differential operators with polynomial coefficients act continuously on this class (i.e. on $\mathcal{S}^\ast$). The Fourier transform on $\mathcal{S}^\ast$ is a topological automorphism.

Any linear differential equation with constant coefficients has a tempered fundamental solution (Hörmander, 1958; Lojasiewicz, 1959). This is a non-trivial refinement of the Ehrenpreis-Malgrange theorem which was formulated above; it is connected with lower bounds for polynomials in several variables, which allow the division problem to be solved successfully in the given situation.

If the basic space $E$ is a contraction of $\mathcal{D}$ with a stronger topology and also $E$ is dense in $\mathcal{D}$, then the generalised functions over $E$ will not now be distributions in general (but all distributions will be generalised functions over $E$). Natural examples of this type of basic space are the non-quasianalytic classes of infinitely differentiable functions (generalised functions over such spaces are given the name ultradistributions). However quasianalytic and especially analytic classes are of no lesser interest in the scheme being discussed. We call continuous linear
functionals on such spaces *analytic functionals*. As a specific case let us consider the space $A(\mathbb{C}^n)$ of entire functions of $n$ complex variables under the topology of uniform convergence on compact sets. In order that a linear functional $\alpha$ on this space be continuous, it is necessary and sufficient that there exist a compact set $Q \subset \mathbb{C}^n$ and a constant $C$ such that

$$|\alpha[f]| \leq C \max_{z \in Q} |f(z)| \quad (f \in \mathcal{F}(\mathbb{C}^n)).$$

We can say that the functional $\alpha$ is *concentrated on* $Q$ but this term is also used in the case where for any neighbourhood $V \ni Q$ there exists a constant $C_V$ such that

$$|\alpha[f]| \leq C_V \sup_{z \in V} |f(z)| \quad (f \in A(\mathbb{C}^n)). \quad (53)$$

**Example.** Let $(\alpha_k)$ be any multi-indexed scalar sequence satisfying the condition

$$\lim_{|k| \to \infty} \frac{1}{k!} |\alpha_k| = 0.$$  

Then\(^{19}\)

$$\alpha[f] = \sum_{k} \alpha_k \left( \frac{\partial^k f}{\partial z^k} \right)_{z=0} \quad (54)$$

is an analytic functional concentrated at 0 and (54) gives the general form of such functionals.

Let us denote by $A^*(Q)$ the set of analytic functionals which are concentrated on the compact set $Q \subset \mathbb{C}^n$. Restricting ourselves now to real compact sets\(^{20}\) (i.e. such that $Q \subset \mathbb{R}^n \subset \mathbb{C}^n$ where the last embedding is the canonical one), let us consider the linear space

$$A'(\mathbb{R}^n) = \bigcup_{r=1}^{\infty} A^*(\{x: x \in \mathbb{R}^n, \|x\| < r\}).$$

Martineau (1961) proved\(^{21}\) that any $\alpha \in A'(\mathbb{R}^n)$ has a *support*, i.e. a smallest compact set $\text{supp} \alpha$ on which $\alpha$ is concentrated. This opened up a relatively simple path to the theory of so-called hyperfunctions on $\mathbb{R}^n$ which was created by Sato (1960). *Hyperfunctions with compact supports* are just the same as functionals from $A'(\mathbb{R}^n)$. If $G$ is a bounded open set in $\mathbb{R}^n$, the space of hyperfunctions on $G$ is by definition the factor space $A^*(\overline{G})/A^*(\partial G)$. The passage to hyperfunctions on the whole of $\mathbb{R}^n$ is effected by means of a covering by bounded neighbourhoods. Thus, roughly speaking, a hyperfunction is a 'locally analytic' functional. On the whole,}

---

\(^{18}\)Thus all continuous linear functionals on $A(\mathbb{C}^n)$ have compact support.

\(^{19}\)Here we use the abbreviation $k! = k \cdot (k-1)! \ldots 1$.

\(^{20}\)But for such compact sets it is still necessary in (53) to consider complex neighbourhoods.

\(^{21}\)In the proof Martineau uses cohomology theory. An 'elementary' proof was given by Hörmander (1983).
this construction is such that hyperfunctions are not generalised functions in the sense used earlier. However in many respects we can work with them just as with generalised (and so also as with classical) functions; in particular we can consider differential equations in hyperfunctions. It is interesting that hyperfunctions were intended first of all to play the role of limiting values of analytic functions in situations where these do not exist in any usual sense. Because of this the setting of hyperfunctions allows us to obtain, for example, adequate generalisations of Bogolyubov's theorem \(\textit{on the edge of the wedge}\) (Martineau, 1968; Morimoto, 1979; Zharinov, 1980). We note that the original version of this theorem was obtained by N.N. Bogolyubov (1956) in connection with the construction of dispersal relations in quantum field theory. The theorem asserts the local analytic continuability of one into the other for a pair of functions \(f_+, f_-\) which are analytic on the domains \(G \pm iK\) (\(G\) is a domain and \(K\) is a cone in \(\mathbb{R}^n\)) under the condition that their limiting values on \(G\) coincide. The extended function turns out to be analytic on some complex neighbourhood \(V \supset G\) (we even have \(V \supset \Gamma\), where \(\Gamma\) is the \(C\)-convex hull of the domain \(G\) (V.S. Vladimirov, 1960)).

5.3. Families of Function Spaces. Although the list of families presented below contains many of the well-known function spaces it is certainly by no means complete and illustrates rather the variety of possible constructions. We note that the nature of a function space is determined to a considerable extent by the structure which is given on the domain of definition \(X\) of those functions from which the space is composed. For example, if \(X\) is a space with a measure, it is natural to consider measurable functions on it; on a topological space \(X\) we consider continuous functions (if we also look at discontinuous functions, they should at least be Borel functions); a metric on \(X\) allows us to introduce the modulus of continuity and to define, for example, Lipschitz classes; on a smooth (analytic) manifold we can consider smooth (analytic) functions and so on.

I. Let \(M(u) (u \geq 0)\) be a convex function such that \(M(u) > 0 (u \neq 0)\) and

\[
\lim_{u \to 0} u M(u) = 0, \quad \lim_{u \to \infty} u M(u) = \infty.
\]

Then we say that this is an \(N\)-function. Its Legendre transform

\[
M^*(v) = \sup_{u \geq 0} (uv - M(u)) \quad (v \geq 0)
\]

is also an \(N\)-function\(^{23}\) and it is said to be dual to \(M\) (it is easy to see that \(M^{**} = M\)).

Let \(X\) be a space with a \(\sigma\)-finite measure \(\mu\). We consider the set \(L_M(X)\) of measurable functions \(v(x) (x \in X)\) which satisfy the condition

\[
M^*(v) \equiv \int_X M^*(|v(x)|) \, d\mu < \infty.
\]

\(^{23}\)i.e. such that the inverse images of Borel sets are of the same type.

\(^{23}\)The non-negativity of \(M^*(v)\) is the classical \(\textit{Young inequality.}\)
We denote by $L^*_M(X)$ the set of functions $u(x)$ such that $uw \in L^1(X, \mu)$ for all $v \in L^*_M(X)$. This linear space has a norm

$$\|u\| = \sup_{M^*(v) \leq 1} \left| \int_X uv \, d\mu \right|$$

(it can be shown that the supremum is finite) under which it turns out to be a Banach space; it is called the Orlicz space with norming function $M$.

**Example.** If $M(u) = \frac{u^p}{p} (p > 1)$, then $M^*(v) = \frac{v^q}{q} \left( \frac{1}{p} + \frac{1}{q} = 1 \right)$ and the Orlicz space $L^*_M$ coincides with $L^p$ (to within a multiple of the norm).

Now let us assume that $M(u)$ is of slow growth in the sense that

$$\sup_{u \in L^*_M} \frac{M(2u)}{M(u)} < \infty.$$ 

The general form of a continuous linear functional on $L^*_M$ is then

$$u \mapsto \int_X uv \, d\mu,$$

where $v \in L^*_M$. By the same token $L^*_M$ turns out to be topologically isomorphic to the conjugate space of $L^*_M$.

II. We associate with each measurable function $u(x) (x \in X)$ its indicatrix

$$m_u(t) = \mu \{ x : |u(x)| > t \} \quad (t \geq 0).$$

This is a non-increasing function which is continuous on the right. Let $\psi(\tau) (\tau \geq 0)$ be a concave, non-decreasing function such that $\psi(0) = 0, \psi(\infty) = \infty$.

The condition

$$\|u\| = \int_0^\infty \psi(m_u(t)) \, dt < \infty$$

defines the Banach space $A_\psi$, which is called a Lorentz space. The conjugate of $A_\psi$ is the Marcinkiewicz space $M_\psi$, which consists of the measurable functions $v(x) (x \in X)$ for which

$$\|v\| = \sup_{A \subseteq \mathcal{A}} \frac{1}{\psi(\mu(A))} \int_A |v| \, d\mu < \infty$$

(the general form of a continuous linear functional on $A_\psi$ is standard).

In the space of measurable functions which are described above it is possible to construct a rich theory of non-linear integral equations, by choosing the space in accordance with the nature of the non-linearity (power, exponential, etc.). This

24 In the probabilistic situation this is a variant of the distribution function of a random variable. Just as we speak of identically distributed random variables in probability theory, here we can speak of equimeasurable functions (i.e. functions whose indicatrices coincide – this is also an important concept).
Chapter 2. Foundations and Methods

programme was realised in the 1950s and 1960s by M.A. Krasnosel'skii in collaboration with several co-authors.

III. Let \( X \) be a metric space with metric \( d \) and let \( u(x) (x \in X) \) be a uniformly continuous function. The quantity

\[
\omega_u(\delta) = \sup_{d(x,y) \leq \delta} |u(x) - u(y)| \quad (\delta > 0)
\]

is called its modulus of continuity. Clearly, \( \omega_u(\delta) \) is a non-decreasing, non-negative function with \( \omega(0_+) = 0 \). Each non-decreasing positive function \( \Omega(\delta) (\delta > 0, \Omega(0_+) = 0) \) defines a space of Lipschitz type\(^{25}\) \( \Omega \), consisting of the continuous functions \( u(x) (x \in X) \) such that

\[
\|u\| = |u(x_0)| + \sup_{\delta > 0} \frac{\omega_u(\delta)}{\Omega(\delta)} < \infty
\]

(\( x_0 \in X \) is a fixed point). This family of Banach spaces serves first and foremost the needs of approximation theory.

IV. Among the spaces of smooth functions let us mention first of all the Sobolev spaces which are defined in their general form by the following procedure. Let \( G \) be an open set in \( \mathbb{R}^n \), let \( 1 < p \leq \infty \) and let \( m \) be an integer \( \geq 1 \). The space \( W_p^m(G) \) consists of those functions whose generalised partial derivatives up to order \( m \) inclusive are functions of class \( \mathbb{L}^p(G) \). Correspondingly we introduce the norm

\[
\|f\|_m = \sum_{|k| \leq m} \left\| \frac{\partial^k f}{\partial x^k} \right\|_{L_p}.
\]

The discrete series of spaces \( W_p^m(G) (m = 1, 2, 3, \ldots) \) is included in the continuous family \( W_p^s(G) (s > 0) \) of Slobodetskij spaces. If \( s = m + r (0 < r < 1) \), it is required of a function \( f \in W_p^s(G) \) that \( f \in W_p^m(G) \) and

\[
\|f\|_s \equiv \|f\|_m + \sum_{|k|=m} \left( \int_G \int_G \left| \frac{\partial^k f}{\partial x^k} (x') - \frac{\partial^k f}{\partial x^k} (x'') \right|^p \frac{dx'}{|x' - x''|^{n+rp}} \right)^{1/p} < \infty.
\]

The first differences of the partial derivatives \( \frac{\partial^k f}{\partial x^k} \) appearing here can be replaced by differences of higher orders\(^{27}\) (this is advisable). The Nikol'skij-Besov spaces, which we now mention, arise in this way in combination with certain other useful modifications. The norms by means of which these spaces are introduced can be replaced by equivalent norms defined in terms of best polynomial approximations in the \( L^p \)-metrics. This fact is the multivariate analogue of the Jackson-Bernshtein theorem.

In the case \( G = \mathbb{R}^n \) the Fourier transform is an effective mechanism for constructing spaces of smooth functions; in terms of it smoothness is regulated by

---

\(^{25}\) For \( \Omega(\delta) = \delta^s (0 < s \leq 1) \) we have the class Lip \( \alpha \).

\(^{26}\) i.e. regular distributions.

\(^{27}\) Here it is necessary to assume that \( G \) is a convex domain.
the rate of decrease at infinity. For a more refined account of smoothness it is appropriate to form a *decomposition* of the Fourier transform with the help of a partition of unity into a sum of functions of compact support \( (\varphi_k)_1^\infty \) in the class \( S \) which are concentrated on a sequence of spherical layers \( r_{k-1} \leq |x| \leq r_k \) \( (r_0 = 0, r_k \to \infty) \):

\[
\hat{f} = \sum_{k=1}^{\infty} \varphi_k \hat{f}.
\]

(55)

Finally, we can require that the sequence of norms (e.g. in \( L^p \)) of the inverse Fourier transforms of the terms in (55) decrease at a predetermined rate in accordance with the growth of the radii \( r_k \). An alternative to the last step is to first calculate a certain norm (taking the radii into account) for the sequence \( (\varphi_k \hat{f})_1^\infty \) and then to require that this function behaves itself sufficiently well at infinity (e.g. belongs to \( L^p \)). These approaches, which were outlined in the mid-1960s (S.M. Nikol'skij, P.I. Lizorkin, Peetre), were systematically developed during the 1970s by Peetre, Triebel and Stöckert and led to the construction of very extensive families of spaces of smooth (of power order) functions which include the Nikol'skij-Besov spaces, the *Lizorkin spaces* associated with derivatives of fractional order and so on. A corresponding theory for the general modulus of continuity has been constructed during the last ten years (G.A. Kalyabin, M.G. Gol'dman).

Function spaces on manifolds are constructed naturally from those on charts (the same can be said about spaces of differential forms and, in general, about sections of vector bundles on manifolds). As has already been said above, in the theory of boundary value problems it is important to know to what space the boundary values of functions from the given space belong. Moreover it is important to have a collection of embedding theorems for individual spaces into others and also to know when the embedding mappings are compact. At the present time extensive information has been accumulated on this account.

### 5.4. Operators on Function Spaces

We have already touched upon this topic repeatedly when speaking of integral and differential operators. A formal notation of the type

\[
f'(x) = -\int_{-\infty}^{\infty} \delta'(x - y) f(y) \, dy
\]

allows us to regard linear differential operators as integral equations with generalised, highly singular kernels. The kernels of ordinary integral operators are regular so that these operators (or their powers) are compact on spaces of type \( L^p \). An intermediate position between the two cases mentioned is occupied by singular integral operators whose kernels are functions with discontinuities of

\footnote{In particular, kernels with bound \( K(x, y) = O(|x - y|^{m-n}) \) \( (m < n) \) on a bounded domain \( G \subset \mathbb{R}^n \) are of this type.}
such a type that, while compactness is lost, we still have boundedness on spaces of type $L^p$.

**Example.** The Hilbert operator

$$
(H\varphi)(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\varphi(\xi) d\xi}{\xi - x} \quad (-\infty < x < \infty)
$$

(56)
is defined and bounded on $L^p(\mathbb{R})$ ($1 < p < \infty$) (theorem of M. Riesz). The integral (56) is obtained as a principal value. If $\varphi \in L^1(\mathbb{R})$, it exists for almost all $x$ (Zygmund, 1959), however the Hilbert operator acts neither on $L^1$ nor on $L^\infty$: if $\varphi(\xi) = 1 (|\xi| \leq 1)$, $\varphi(\xi) = 0 (|\xi| > 1)$, then $(H\varphi)(x) = \ln \left| \frac{x + 1}{x - 1} \right|$. It is useful for the investigation of the operator $H$ to pass to the Fourier transforms $\hat{\varphi}$, $\hat{H\varphi}$. Then it transforms into the operator of multiplication by $\frac{1}{2} \text{sgn} \lambda$. However the Fourier transform itself acts on $L^p$ only for $p = 2$, so that for the time being only boundedness of the Hilbert operator on $L^2$ is clear. From certain considerations of a special character we can obtain in addition its boundedness for $p = 4, 6, \ldots$, after which the passage to arbitrary $p \geq 2$ is effected with the help of the following.

**M. Riesz-Thorin Interpolation Theorem.** Let $1 \leq p_i \leq r_i \leq \infty$ ($i = 0, 1$) with $p_i$ finite, put $p_x = [(1 - \alpha)p_0^{-1} + \alpha p_1^{-1}]^{-1}$ ($0 \leq \alpha \leq 1$) and let $r_x$ be defined similarly.

If an operator $A$ is defined on the set of piecewise-constant functions $\varphi(\xi)$ of compact support ($-\infty < \xi < \infty$) and maps them into functions in $L^{r_0} \cap L^{r_1}$ in such a way that $\|A\varphi\|_{r_i} \leq M_i \|\varphi\|_{p_i}$ ($i = 0, 1$), then

$$
\|A\varphi\|_{r_x} \leq M_x \|\varphi\|_{p_x} \quad (0 \leq \alpha \leq 1), \quad \text{where} \quad M_x = M_{\alpha}^{r_0}M_{1-x}^{r_1}
$$

(|| \cdot ||_x is the norm on $L^x$).

This theorem together with its immediate generalisations (Marcinkiewicz, 1939) became the starting point of the general theory of interpolation of linear operators which was referred to above. We will describe briefly the subject of this theory which at the present time is already far advanced and finds diverse applications.

Let $E_0, E_1$ be a pair of Banach spaces both of which are continuously embedded in the same LTS (a Banach pair). We can then construct the Banach spaces $E_0 \cap E_1, E_0 + E_1$ by putting

$$
\|x\|_{E_0 \cap E_1} = \max(\|x\|_{E_0}, \|x\|_{E_1}),
$$

$$
\|x\|_{E_0 + E_1} = \inf_{x = x_0 + x_1} (\|x_0\|_{E_0} + \|x_1\|_{E_1}).
$$

A Banach space $E$ which is contained in $E_0 + E_1$ and contains $E_0 \cap E_1$ is said to be intermediate for the Banach pair $(E_0, E_1)$ if the embeddings $E_0 \cap E_1 \rightarrow$
4 5. Function Spaces

\[ E \rightarrow E_0 + E_1 \] are continuous. In particular, \( E_0 \) and \( E_1 \) are themselves of this type. We also note that, if \( E_0 \subset E_1 \) and the embedding is continuous, then \( E_0 \cap E_1 \approx E_0 \) and \( E_0 + E_1 \approx E_1 \).

An intermediate space for a Banach pair \((E_0, E_1)\) is said to be an interpolation space if it is invariant for all linear operators \( A \) on \( E_0 + E_1 \) for which both \( E_0 \) and \( E_1 \) are invariant and the operators \( A|_{E_0} \), \( A|_{E_1} \) are bounded for the corresponding norms\(^3\) (the linear space of such operators is denoted by \( I(E_0, E_1) \)).

The Interpolation Inequality. If \( E \) is an interpolation space for a Banach pair \((E_0, E_1)\), then \( A|_E \) is bounded and

\[ \|A|_E\| \leq M_E \max(\|A|_{E_0}\|, \|A|_{E_1}\|), \quad (57) \]

where \( M_E \) is a constant independent of \( A \) (interpolation constant)\(^3\).

**Proof.** The operator \( A|_E \) is closed, since it is closed (even bounded) under the weaker norm induced from \( E_0 + E_1 \). It is therefore bounded. Further the space \( I(E_0, E_1) \) is a Banach space with respect to both the norm which appears on the right hand side in (57) and the weaker norm \( \|A\|_{E_0 + E_1} \). These norms are therefore equivalent. Finally, the restriction homomorphism of \( I(E_0, E_1) \to \mathcal{L}(E) \) is closed (which is easily verified with the norm \( \|A\|_{E_0 + E_1} \) on \( I(E_0, E_1) \)). It is therefore bounded, i.e. (57) holds. \( \square \)

**Remark.** In the more general situation to which, in particular, the Riesz-Thorin theorem relates we have to consider two Banach pairs \((E_0, E_1), (F_0, F_1)\), spaces \( E, F \) which are intermediate for these pairs, and an interpolation in the sense that, for all homomorphisms \( A: E_0 + E_1 \to F_0 + F_1 \) such that \( AE_0 \subset F_0 \), \( AE_1 \subset F_1 \) and \( A|_{E_0}, A|_{E_1} \) are bounded, we have \( AE \subset F \). There is no problem in carrying over the interpolation inequality to this situation.

As an example let us consider the Banach pair \((L^1, L^\infty)\). The role of the embedding \( L^1 \) is taken here by the space \( \Phi \) of measurable functions with the topology of convergence in measure on each finite interval. The space \( \Phi_0 = L^1 + L^\infty \) is topologically isomorphic to the space of functions \( \varphi \in \Phi \) for which the indicatrix \( m_\varphi \neq \infty \). A Banach space \( E \) which is continuously embedded in \( \Phi_0 \) is said to be symmetric if 1) whenever \( \varphi \in E, \psi \in \Phi_0 \) and \( |\psi(\xi)| \leq |\varphi(\xi)| \) almost everywhere, it follows that \( \psi \in E \) and \( \|\psi\|_E \leq \|\varphi\|_E \), 2) whenever \( \varphi \in E, \psi \in \Phi_0 \) and \( \varphi, \psi \) are equimeasurable, it follows that \( \psi \in E \) and \( \|\psi\|_E = \|\varphi\|_E \). For example, all \( L^p \) \((1 \leq p \leq \infty)\) are of this type, as are Orlicz, Lorentz and Marcinkiewicz spaces. E.M. Semenov (1964) proved that all symmetric spaces are intermediate for \((L^1, L^\infty)\) and any interpolation space for a pair of symmetric spaces is topologically isomorphic to a symmetric space (it is symmetric if the interpolation constant is equal to 1). Calderón (1966) obtained a complete description of interpolation spaces for \((L^1, L^\infty)\). In any event all separable sym-

\footnote{Under these conditions \( A \) is bounded both on \( E_0 + E_1 \) and on \( E_0 \cap E_1 \) for the norms introduced on these spaces.}

\footnote{It can at once be assumed to be the least such constant. If we then have \( M_E = 1 \), the interpolation space is said to be exact.}
metric spaces are interpolation spaces for \((L^1, L^\infty)\). The particular symmetric spaces mentioned above are interpolation spaces for \((L^1, L^\infty)\). Nevertheless there does exist a symmetric space which is not an interpolation space for the pair \((L^1, L^\infty)\) (G.I. Russu, 1969).

In situations where it does not seem to be possible to describe all interpolation spaces for some Banach pair, a sensible approach is to find an effective construction for producing a sufficiently extensive family of interpolation spaces. The interpolation methods\(^3\) proposed for this purpose are functorial constructions and it is precisely in this that the guarantee of their success is contained. This point of view, which goes back to work of Aronszajn and Gagliardo (1965), has recently taken on a fundamental character (Yu.A. Brudnyj, N. Kruglyak).

One further significant aspect of the theory of interpolation spaces is connected with one-parameter families of Banach spaces \(E_\alpha (0 \leq \alpha \leq 1)\) having the following properties:

1) for \(\beta > \alpha\) the space \(E_\beta\) is continuously and densely embedded in \(E_\alpha\);
2) for \(\gamma > \beta > \alpha\)

\[
\|x\|_{E_\gamma} \leq C(\alpha, \beta, \gamma) \|x\|_{E_\beta}^{(\gamma-\beta)/(\gamma-\alpha)} \|x\|_{E_\alpha}^{(\beta-\alpha)/(\gamma-\alpha)} \quad (x \in E_\gamma),
\]

where \(C(\alpha, \beta, \gamma)\) is a positive constant. Such a family is said to be a scale (normal if \(\|x\|_{E_\alpha} \leq \|x\|_{E_\beta}\) \((\beta > \alpha)\) and \(C(\alpha, \beta, \gamma) \equiv 1\)). Many scales \((E_\alpha)_{0 \leq \alpha \leq 1}\) constructed under specific conditions turn out to consist of interpolation spaces for the pair \((E_0, E_1)\) (Lions, 1958; S.G. Krejn, 1960; S.G. Krejn - Yu.I. Petunin, 1966)\(^33\).

Because of this occurrence, scales prove to be a suitable tool for investigating boundary value problems for elliptic differential operators (Magenes, 1963).

Returning to the consideration of specific operators on function spaces, we note first of all the multivariable analogue of the Hilbert operator:

\[
(H_\alpha \varphi)(x) = \int_{\mathbb{R}^n} \frac{\Omega(x, \theta)}{|y|^n} \varphi(x - y) \, dy.
\]  

(58)

Here \(\varphi_y = y/|y|\), the function \(\Omega(x, \theta)\) is continuous (for simplicity) and

\[
\int_{|\theta|=1} \Omega(x, \theta) \, d\theta = 0.
\]

The integral (58) is obtained as a principal value. The singular integral operator \(H_\alpha\) is defined and bounded on \(L^p(\mathbb{R}^n)\) \((1 < p < \infty)\) if \(\Omega\) belongs to \(L^q\) with respect to \(\theta \left(\frac{1}{p} + \frac{1}{q} = 1\right)\) and its \(L^p\)-norm is bounded with respect to \(x\) (the case \(n = 2, p = 2\) is due to S.G. Mikhlin, 1953; the general case is due to Calderón and Zygmund, 1956).

\(^3\)Real' and 'complex' methods differ here. The latter are based on results from complex analysis such as Hadamard's three-lines theorem. We note that with these methods it is possible to show, for example, that Nikol'skii-Besov spaces are interpolation spaces for a pair of spaces from this same family or for a pair of Sobolev spaces etc.

\(^33\)In particular, the scales which can be constructed by the complex interpolation method.
The Fourier transform with respect to $y$ of the kernel of the operator $H_\Omega$ is called the symbol of this operator:

$$\sigma(x, \xi) = \int_{\mathbb{R}^n} \frac{\Omega(x, \theta_y)}{|y|^n} e^{-i\xi \cdot y} \, dy.$$ 

If $\Omega$ and consequently also $\sigma$ do not depend on $x$, then

$$(H_\Omega \varphi)(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \sigma(\xi) \phi(\xi) e^{i\xi \cdot x} \, d\xi.$$ 

In this case the composition theorem holds: the product of the symbols corresponds to the product of the operators (Calderón-Zygmund, 1952). This result has a natural connection with the convolution theorem but it is not so obvious since the integrals concerned are principal values.\(^{34}\) Since the correspondence between operators and symbols is clearly also linear, we have here a homomorphism of an algebra of operators into an algebra of functions, i.e. a certain variant of operator calculus. This is an effective route to the solution of multivariate singular integral equations (this was in fact projected in the late 1920s by Tricomi).

In the general case it is useful to consider along with $H_\Omega$ the operator $G_\Omega$, whose investigation can turn out to be simpler than that of $H_\Omega$, will at the same time imitate $H_\Omega$ sufficiently well. The operator $G_\Omega$ is pseudodifferential. To explain this term let us consider an operator of the form

$$(P \varphi)(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} p(x, \xi) \phi(\xi) e^{i\xi \cdot x} \, d\xi,$$

where $p$, an arbitrary given function of power growth rate\(^{35}\) with respect to $\xi$, is the symbol of the operator $P$. If $p(x, \xi)$ is a polynomial in $\xi$ (with coefficients depending on $x$), then $P$ is a differential operator (with the same coefficients). If we eliminate the intermediate Fourier transform, we can represent the operator $P$ as a double integral:

$$(P \varphi)(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} p(x, \xi, \eta) \varphi(y) e^{i(\xi, x-y)} \, d\xi \, dy. \quad (59)$$

---

\(^{34}\)The concept of the symbol was introduced differently at an early stage in the development of the theory of singular integral equations and correspondingly the composition theorem was proved in another context (S.G. Mikhlin, 1936; Giraud, 1936; S.G. Mikhlin, 1955). Then the equivalence of the two definitions of the symbol was established (S.G. Mikhlin, 1956).

\(^{35}\)Uniformly with respect to $x$ on each compact set. The exponent of the power of growth (not necessarily integral) is called the order of the operator. Usually $p$ is a function of class $C^\infty$ with respect to all its variables with suitable bounds on the derivatives.
The final step leading to the concept of a pseudodifferential operator (PDO) in full generality consists of replacing $p(x, \xi)$ in formula (59) by $p(x, y, \xi)$ (it now follows that we should call an operator of the form (59) a PDO in the restricted or proper sense). This class of operators is sufficient for studying many questions in the theory of elliptic differential equations (including the solution of the index problem (Atiyah-Singer, 1968), in which PDOs played an important role). We can regard the construction of a PDO as a quantization which associates with each scalar function (symbol) an operator on a suitable function space. Thus, for example, problems such as quasiclassical asymptotics (passage to the limit from quantum mechanics to the classical situation by letting Planck’s constant $\hbar$ tend to zero) are naturally studied in the setting of PDOs and their further generalisations (V.P. Maslov, 1965; V.P. Maslov, M.V. Fedoryuk, 1976; V.L. Rojtburd, 1976). The generalisations mentioned are Fourier integral operators (FIOs), whose theory was formulated during the mid 1960s and early 1970s thanks to works of V.P. Maslov, Hörmander and other authors. PDOs were already in use at the beginning of this period, however many problems (including quasiclassical asymptotics and the Cauchy problem for hyperbolic equations) required the introduction of FIOs. The operators are constructed on Lagrange manifolds by gluing together local FIOs which have a similar form to PDOs but with exponential term $\exp(iS(x, y, \xi))$, where in general $S$ is some non-linear function.

In conclusion we note that in the development of functional analysis functions of infinitely many variables played and continue to play a significant role. Such, for example, are the functionals which arise in variational calculus, states with an unrestricted number of particles in quantum field theory, and so on. Rigorous analysis has been successfully constructed on spaces of functions of infinitely many variables. A systematic account of it is contained in the monographs by Yu.M. Berezanskij (1978) and by Yu.L. Daletskij and S.V. Fomin (1983). It is most important here to have a satisfactory theory of measure and integration (in the field of applications we point out the Wiener measure in the theory of random processes, the Feynman integral in quantum mechanics and electrodynamics, secondary quantization). A fundamental difficulty is caused by the lack of a translation invariant measure on an infinite-dimensional LTS since it is not locally compact (the last condition is close to being necessary in the theorem of Haar). However for many purposes the so-called quasi-invariance is sufficient. In particular, Gaussian measures’, which include the Wiener measure, have this property. It is clear from what has been said that on an infinite-dimensional space we have to construct not measures ‘general’ but special classes of ‘good’ measures. A certain axiomatic scheme for measure theory in an infinite-dimensional space was proposed by A.M. Vershik (1968) and he and V.N. Sudakov (1969) conducted an extensive investigation of probability measures. Generally speaking the connections between functional analysis and probability theory are deep and varied. A meaningful account of this aspect lies beyond the scope of the present volume.
Commentary on the Bibliography

The principal events in the foundation of functional analysis have been elucidated in the articles of Bourbaki and in the book by Dieudonné (1983). This discipline attained a definite maturity in 1932, which is testified to by the simultaneous appearance in this year of three fundamental works: "Théorie des opérations linéaires" by Banach, "Mathematische Grundlagen der Quantenmechanik" by von Neumann and "Linear transformations in Hilbert space and their applications to analysis" by Stone. However, many important topics (linear topological spaces, ordered linear spaces, Banach algebras etc.) were developed later on and surveys and monographs devoted to them appeared correspondingly later. At the present time, there are many hundreds of titles of individual books dealing just with functional analysis and its applications. We hope that the selection presented below is representative of the subject (but it is by no means representative in terms of priority). As far as articles are concerned, the list includes only certain surveys from "Sbepkhi Matematicheskikh Nauk" (in order to supplement the monographs) and works from previous volumes in the present series which are connected with our subject. Several books are also mentioned from the "Spravochnaya Matematicheskaya Biblioteka", including the publication "Functional Analysis" (1972) which was produced by 17 authors under the general editorship of S.G. Krejn.

References in the text to articles (author, year) are not associated with the Bibliography. In the majority of cases, the original source will be found without difficulty from the given information.

Bibliography*

I. Basic Texts


II. Additional Educational Literature


*For the convenience of the reader, references to reviews in Zentralblatt für Mathematik (Zbl.), compiled using the MATH database, have, as far as possible, been included in this bibliography.


Halmos, P.R. [1956]: Lectures on Ergodic Theory. Math. Soc. Japan: Tokyo, Zbl.73,93


Bibliography

Sadovnichii, V.A. [1986]: Theory of Linear Operators. Moscow Univ. Press: Moscow (Russian)
Trenogin, V.A. [1980]: Functional Analysis. Nauka: Moscow (Russian), Zbl.571.46001

III. Monographs

Azizov, T.Ya., Iokhvidov, I.S. [1986]: Basic Theory of Linear Operators on Spaces with Indefinite Metric. Nauka (Russian), Zbl.607.47031


Levitan, B.M. [1953]: Almost-periodic Functions. State Science Press: Moscow (Russian)
Maak, W. [1950]: Fastperiodische Funktionen. Springer-Verlag: Berlin, Zbl.37,57
Mackey, G.W. [1965]: Group Representations and Commutative Harmonic Analysis with Applications to Analysis, Number Theory and Physics. Univ. Press: Berkeley


Nakano, H. [1950]: Modularized Semi-ordered Linear Spaces. Mazuren: Tokyo


Peczynski, A. [1968]: Linear Extensions, Linear Averagings and their Applications to Linear Topological Classification of Spaces of Continuous Functions. Diss. Math.: Warszawa, Zbl.165,146


Pietsch, A. [1978]: Operator Ideals. VEB Deutscher Verlag der Wiss.: Berlin, Zbl.399,47039


Rudin, W. [1980]: Function Theory in the Unit Ball of C^n. Springer-Verlag: Berlin, Zbl.495,32001


Bibliography

Suetin, P.K. [1984]: Series of Faber Polynomials. Nauka: Moscow (Russian)
Wiener, N. [1933]: The Fourier Integral and Certain of its Applications. Cambridge Univ. Press, Zbl.6,54
Zelazko, W. [1984]: On Ideal Theory in Banach and Topological Algebras. Monogr. Inst. Mat. Universidad Nacional Autónoma de Mexico, No. 15: Mexico City

IV. Survey Articles from Usp. Mat. Nauk

Bogolyubov, N.N. [1965]: Mathematical problems of quantum field theory. 20, No. 3, 31–40 (Russian)
Fet, A.I. [1956]: A space of analytic functions and its application to the Cauchy-Kovalevski problem. 11, No. 2, 215–222 (Russian), Zbl.70,88
Gel'fand, I.M. [1956]: On certain problems of functional analysis. 11, No. 6, 3–12, Zbl.73,94. English transl.: Am. Math. Soc. Transl., II. Ser. 16, 315–324 (1960)

Kantorovich, L.V. [1948]: Functional analysis and applied mathematics. 3, No. 6, 89–185 (Russian), Zbl.34,212


Kantorovich, L.V. [1948]: Functional analysis and applied mathematics. 3, No. 6, 89–185 (Russian), Zbl.34,212


Krein, M.G., Krasnosel’skii, M.A. [1947]: Fundamental theorems on the expansion of linear operators and some of their applications to the theory of orthogonal polynomials and the moment problem. 2, No. 3, 60–160 (Russian)

Krejn, M.G., Rutman, M.A. [1948]: Linear operators leaving invariant a cone in a Banach space. 3, No. 1, 3–95 (Russian), Zbl.30,129


Peller, V.V., and Krushchev, S.V. [1982]: Hankel operators, best approximations and stationary

V. Articles from Previous Volumes of the Encyclopaedia of Math. Sciences


VI. Reference Books from the Series Spravochnaya Matematicheskaya Biblioteka

Babich, V.M. et al. [1964]: Linear Equations of Mathematical Physics. (Mikhlin, S.G. (ed.)) Nauka: Moscow, Zbl.122,330


Zabrejko, P.P. et al. [1968]: Integral Equations. Nauka: Moscow
Author Index

Adamyan, V.M. 227
Ahiezer, N.I. 29, 61
Alakhverdiev, D.Eh. 216
Alaoglu, L. 145
Aleksandrov, P.S. 175
Amir, D. 143
Appel, P.E. 65
Arens, R.F. 148, 236
Aronszajn, N. 175, 258
Arov, D.Z. 227
Arzelá, C. 154
Ascoli, G. 113, 154
Atiyah, M.F. 95, 260
Atkinson, F.V. 162, 167

Babenko, K.I. 135
Babich, V.M. 244
Bade, W.G. 207
Baire, R. 122
Bari, N.K. 159, 169
Berezanskij, Yu.M. 203, 242, 260
Bernoulli, D. 19
Bernoulli, Johann 10
Bernshtein, S.N. 26, 29, 30, 43, 55, 57, 63
Besov, O.V. 243
Bessel, F.W. 24, 28, 52
Beurling, A. 226
Birkhoff, G.A. 117
Birman, M.S. 206
Bishop, E. 130, 209, 238
Blaschke, W. 68, 224
Bochkarev, S.V. 134
Bochner, S. 35, 52, 53, 141, 241, 249
Bogolyubov, N.N. 179, 246, 252
Bohl, P. 191
Bohr, H. 51, 52, 191
Bonsall, F.F. 215
Borel, E. 53, 69, 151
Borok, V.M. 5, 249
Bourbaki, N. 261
Brauer, A. 174
Brodzki, M.S. 221–223
Brudnyj, Yu.A. 5, 258
Bunyakovskij, V.Ya. 13
Burnside, W. 177, 180

Calderón, A.P. 236, 243, 257–259
Calkin, J.V. 217

Carathéodory, K. 108
Carleman, T. 30, 70
Carleson, L. 23, 214, 233
Cauchy, A. 12, 13, 144
Chebyshev, P.L. 53, 54, 58, 59
Choquet, G. 143
Clairault, A.C. 20
Clarkson, J.A. 130, 158
Cotes, R. 64
Courant, R. 47

D'Alembert, J. 20
Daletskij, Yu.L. 260
Day, M.M. 131, 158
De Leeuw, K. 192
Delsarte, J. 44, 204
Denjoy, A. 26
Dieudonné, J. 152, 261
Dirac, P.A.M. 11
Dirichlet, P.G.L. 17, 22, 42
Dolzhenko, E.P. 66
Du Bois-Reymond, P. 23
Duncan, J. 215
Dunford, N. 143, 207, 209
Dvoretzki, A. 136, 172
Dynin, A.S. 172

Eberlein, W.F. 151
Efimov, N.V. 147
Ehrenpreis, L. 144
Ehrenpreis, S.D. 249
Eidelheit, M. 113
Enflo, P. 134, 158, 166
Esclangon, E. 93
Euclid 4
Euler, L. 10, 19, 20

Faber, G. 63, 66
Faddeev, L.D. 205, 227
Fedoryuk, M.V. 260
Fejér, L. 23, 53, 117
Fekete, M. 184
Feller, W. 189
Fermat, P. 10
Fischer, E. 14, 47
Fixman, V. 207
Foiias, C. 211, 214, 215, 225, 226
Fomin, S.V. 260
Fourier, J. 19, 20, 50
Fréchet, M. 191
Fredholm, I. 74, 165, 166
Frobenius, G. 175, 180, 228
Fubini, G. 15
Gagliardo, E. 243, 258
Gakhov, F.D. 82
Gelfand, I.M. 27, 48, 107, 140, 190, 204, 232, 233, 239–241, 249
Gelfand, A.O. 62
Ginzburg, Yu.P. 229
Giraud, G. 259
Glazman, I.M. 213
Glicksberg, I. 192
Gindikin, K. 97
Gokhberg, I.Ts. 138, 162, 167, 221
Gol'den, V.M. 255
Gordine, A. 5, 190, 235
Grothendieck, A. 152, 170–173
Haar, A. 56, 179
Hadamard, J. 10, 69, 144, 258
Hahn, H. 112
Hamburger, H.L. 30
Hartman, P. 51
Hartogs, F. 66
Hausdorff, F. 34, 154, 215
Heaviside, O. 40
Helly, E. 31, 193
Helly, E. 35, 145
Herglotz, G. 35
Hermite, C. 62
Hilbert, D. 14, 27, 74, 103, 105, 187, 193, 201
Hille, E. 189
Holub, J.R. 166
Hopf, E. 79
Hopf, H. 175
Hörmander, L. 248, 250, 260
Huff, R.E. 143
Hunt, R.A. 23
Iokhvidov, I.S. 228, 230
Jackson, D. 55, 57
Jacobs, K. 192
James, R.C. 130, 147, 158
Jordan, G. 22, 97, 98, 224
Jordan, P. 130
Kadets, M.I. 5, 122, 135, 158, 191
Kadets, V.M. 5, 136
Kakutani, S. 121, 177, 192
Kalish, G.K. 223
Kalyabin, G.A. 255
Kantorovich, L.V. 117, 163
Karlin, S. 135
Kato, T. 206
Katsnelson, V.Eh. 5, 190, 214, 215
Keldysh, M.V. 65, 217, 231
Kheifits, A.I. 133
Khinchin, A.Ya. 35
Klee, V.L. 131
Koch, H. 171
Kolmogorov, A.N. 12, 23, 56, 93, 100, 107, 122, 126
König, H. 219
Koopman, D.O. 191
Korkin, A.N. 58
Kostyuchenko, A.G. 48
Kotelnikov, V.A. 69
Krasnosel'skiy, M.A. 137, 182, 254
Krejn, M.G. 28, 61, 79, 115, 129, 137–139, 146, 153, 162, 169, 182, 204, 206, 221, 228, 230, 241
Krejn, S.G. 189, 242, 243, 258, 261
Kruglyak, N. 258
KruLL, W. 101
Krylov, N.M. 179
Kupedja, S.T. 206
Kuzmin, R.O. 64
Ladyzhenskaya, O.A. 249
Lagrange, J.L. 10, 17
Lang, S. 105
Langer, H. 214, 230
Lavrent'ev, M.A. 65
Lax, P.D. 227
Lebesgue, H. 12, 15, 23, 55, 151
Leibniz, G.W. 6
Leray, J. 165
Levin, B.Ya. 29, 56, 205
Levinson, N. 135
Levitan, B.M. 48, 204
Levy, P. 80, 235
Lidskij, V.B. 219, 220
Lin, V.Ya. 235
Lindemann, F. 62
Lindenstrauss, J. 143, 157
Lions, J.L. 243, 258
Liouville, J. 46, 47, 184
Litvinov, G.M. 170, 171
Livshits, M.S. 220, 223–225, 227
Lizorkin, P.I. 255
Lobachevskij, N.I. 20
Lojasiewicz, S. 250
Lomonosov, V.I. 176, 210
Lorch, E.R. 191
Author Index

Lumer, G. 215
Luzin, N.N. 23, 26
Lyubarskij, Yu.I. 5
Lyubich, M.Yu. 5, 192
Lyubich, Yu.I. 43, 44, 92, 188, 192, 209, 211

Mackey, G.W. 148, 149
Magenes, E. 258
Mairhuber, J.C. 56
Malgrange, B. 247
Marichev, A.I. 97
Marchenko, V.A. 204, 205
Marcinkiewicz, J. 256
Markov, A.A. 57–60
Markov, A.A. (junior) 177
Markus, A.S. 5, 138, 170, 131, 146, 233
Markushevich, A.I. 135
Markov, A.A. 57–60
Matsaev, V.I. 170, 205, 209, 210, 215, 217, 219, 222
Mautner, F.I. 203
Mazur, S. 107, 113, 122, 131, 146, 233
Mellin, H. 39
Meyer, T. 78
Mergelyan, S.N. 29, 65, 66
Mikhlin, S.G. 168, 258, 259
Milman, D.P. 130, 137–139, 162, 182
Milyutin, A.I. 122
Minkowski, H. 14, 138
Mityagin, B.S. 172
Miyadera, I. 189
Molchanov, A.M. 51
Molin, F.Eh. 180
Montel, P. 151
Morimoto, M. 252
Morris, P. 143
Mukminov, B.R. 213
Müntz, C.H. 133

Najmark, M.A. 226, 239, 241
Nevanlinna, R. 31
Newton, I. 6
Nikol’skii, O. 142
Nikolaev, L.N. 210
Nikolskij, N.K. 5, 210, 212, 220, 221, 227
Nikol’skii, S.M. 243, 255
Noether, E. 89
Noether, M. 82
Numakura, K. 192

Ostrowski, A.M. 70, 210

Paley, R.E.A.C. 38, 163
Parasyuk, O.S. 246
Parseval, M. 25, 28, 37, 46, 48, 49, 53, 135
Pastur, L.A. 5
Pavlov, B.S. 227
Peetre, J. 255
Pełczyński, A. 135
Perron, O. 175
Peter, F. 180, 241
Petrovkij, I.G. 248, 249
Petit, B.J. 140, 143
Petunin, Yu.I. 258
Phelps, R.R. 130, 143
Phillips, R.S. 189, 215, 227, 229
Picard, E. 75, 164
Pietsch, A. 169, 172
PlancheL, M. 37, 50, 241
Pontryagin, L.S. 228, 230, 241
Potapov, V.G. 224
Povzner, A.Ya. 203, 204
Pták, V. 152
Pythagoras 13, 25

Radon, J. 142
Rajkov, D.A. 149, 241
Read, C.J. 175
Riemann, B. 12, 23, 39, 136
Riesz, F. 14, 16, 35, 116, 117, 129, 131, 132, 165, 208
Riesz, M. 31, 256
Ringrose, J.R. 220
Roberts, J.W. 139
Rogers, C.A. 136, 172
Rojtburd, V.L. 260
Rosenblum, M. 206
Rosenthal, A. 66
Royden, H. 236
Rudin, W. 212
Runge, K. 65
Russu, G.I. 258
Rutman, M.A. 146, 162, 175, 186

Sakhnovich, L.A. 204, 222
Schauder, J. 167, 174
Schmidt, E. 74
Schoenberg, I. 57
Schrödinger, E. 46, 47, 203
Sohur, I. 177, 180, 214
Schwartz, H.A. 13
Schwartz, L. 45, 242, 248, 249
Semenov, E.M. 257
Shilov, G.E. 235–237, 249, 250
Shmul’yan, V.L. 130, 139, 151–153, 158
Shmul’yan, Yu.L. 225
Shtraus, A.V. 225
<table>
<thead>
<tr>
<th>Author</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sieklucki, K.</td>
<td>56</td>
</tr>
<tr>
<td>Singer, I.M.</td>
<td>95, 260</td>
</tr>
<tr>
<td>Slobodetskij, L.N.</td>
<td>244</td>
</tr>
<tr>
<td>Smith, K.T.</td>
<td>175</td>
</tr>
<tr>
<td>Sobolev, S.L.</td>
<td>242–244</td>
</tr>
<tr>
<td>Sobolevskij, P.E.</td>
<td>189</td>
</tr>
<tr>
<td>Sokhotskij, Yu.V.</td>
<td>83</td>
</tr>
<tr>
<td>Stechkin, S.B.</td>
<td>147</td>
</tr>
<tr>
<td>Steinhaus, H.</td>
<td>123</td>
</tr>
<tr>
<td>Steinitz, E.</td>
<td>136</td>
</tr>
<tr>
<td>Steklov, V.A.</td>
<td>47</td>
</tr>
<tr>
<td>Stieltjes, K.</td>
<td>12, 28</td>
</tr>
<tr>
<td>Stöckert, B.</td>
<td>255</td>
</tr>
<tr>
<td>Stone, M.H.</td>
<td>104, 201, 233, 234, 261</td>
</tr>
<tr>
<td>Sturm, C.</td>
<td>46, 47</td>
</tr>
<tr>
<td>Sudakov, V.N.</td>
<td>260</td>
</tr>
<tr>
<td>Sushkevich, A.K.</td>
<td>192</td>
</tr>
<tr>
<td>Sylvester, J.J.</td>
<td>99</td>
</tr>
<tr>
<td>Szökefalvi-Nagy, B.</td>
<td>214, 225, 226</td>
</tr>
<tr>
<td>Täkling, S.</td>
<td>243</td>
</tr>
<tr>
<td>Tannaka, T.</td>
<td>241</td>
</tr>
<tr>
<td>Thorin, G.O.</td>
<td>256</td>
</tr>
<tr>
<td>Tikhonov, A.N.</td>
<td>145, 174, 243</td>
</tr>
<tr>
<td>Titchmarsh, E.C.</td>
<td>45, 48, 223</td>
</tr>
<tr>
<td>Tkachenko, V.A.</td>
<td>5, 44, 209</td>
</tr>
<tr>
<td>Toeplitz, O.</td>
<td>96, 123, 159, 193</td>
</tr>
<tr>
<td>Tonelli, L.</td>
<td>56</td>
</tr>
<tr>
<td>Tricomi, F.G.</td>
<td>259</td>
</tr>
<tr>
<td>Triebel, H.</td>
<td>255</td>
</tr>
<tr>
<td>Tzafriri, L.</td>
<td>157</td>
</tr>
<tr>
<td>Van Kampen, E.R.</td>
<td>241</td>
</tr>
<tr>
<td>Vershik, A.M.</td>
<td>5, 260</td>
</tr>
<tr>
<td>Vitushkin, A.G.</td>
<td>65</td>
</tr>
<tr>
<td>Vladimirov, V.S.</td>
<td>252</td>
</tr>
<tr>
<td>Volterra, V.</td>
<td>10</td>
</tr>
<tr>
<td>Von Neumann, J.</td>
<td>124, 130, 178, 179, 191, 193, 194, 202, 214, 215, 261</td>
</tr>
<tr>
<td>Walsh, J.L.</td>
<td>65</td>
</tr>
<tr>
<td>Weierstrass, K.</td>
<td>24, 28, 65, 68, 151, 233, 234</td>
</tr>
<tr>
<td>Weil, A.</td>
<td>241</td>
</tr>
<tr>
<td>Wermer, J.</td>
<td>220, 234</td>
</tr>
<tr>
<td>Weyl, H.</td>
<td>31, 48, 50, 51, 180, 202</td>
</tr>
<tr>
<td>Widder, D.V.</td>
<td>40,189</td>
</tr>
<tr>
<td>Wiener, N.</td>
<td>27, 38, 39, 79, 80, 163, 235, 236</td>
</tr>
<tr>
<td>Williamson, J.H.</td>
<td>165</td>
</tr>
<tr>
<td>Wintner, A.</td>
<td>51</td>
</tr>
<tr>
<td>Wold, H.</td>
<td>214</td>
</tr>
<tr>
<td>Yang, C.T.</td>
<td>57</td>
</tr>
<tr>
<td>Yentsch, R.</td>
<td>175</td>
</tr>
<tr>
<td>Yosida, K.</td>
<td>189, 192</td>
</tr>
<tr>
<td>Young, L.C.</td>
<td>252</td>
</tr>
<tr>
<td>Zharinov, V.V.</td>
<td>252</td>
</tr>
<tr>
<td>Zobin, N.M.</td>
<td>172</td>
</tr>
<tr>
<td>Zolotarev, E.I.</td>
<td>58</td>
</tr>
<tr>
<td>Zolotarev, G.N.</td>
<td>249</td>
</tr>
<tr>
<td>Zygmund, A.</td>
<td>256, 259</td>
</tr>
</tbody>
</table>
Subject Index

Action 10, 178
Affine manifold 88
Algebra, Banach 27, 232
  — commutative 27
  — Boolean 28
  — function 101
  — group 240
  — local 101
  — of sets 12
  — operator 239
  — reducible 176
  — semisimple 102
  — spectral 102
  — symmetric 234
  — uniform 234
  — universally enveloping 242
  — Wiener 27
  — with weight 237
  — with involution 238
Algebraic interior 108
Almost-period 52
Annihilator of a subspace 89

Ball, q-unit 111
Banach limit 178
Barrel 122
Bases, quadratically close 169
Basis, Bari 169
  — Hamel 85
  — Riesz 159
  — Schauder 134
  — unconditional 135
C*-algebra 239
Capacity, analytic 65
  — continuous 65
Central symmetry 108
Centralizer of a subalgebra 176
Centre of an algebra 176
  — of mass 143
Character of a representation 180
  — of a semigroup 180
  — of an algebra 104
Charge 131
Chebyshev alternant 53
Class, Hardy 38, 225
  — Hölder 26
  — Lipschitz 26
  — Nevanlinna 32
  — quasianalytic 69
Closure of a linear operator 193
Codefect of a homomorphism 90
Codimension of a subspace 88
Coefficients, Christoffel 60
  — Fourier 21, 52
  — generalised 27
Coincidence of a homomorphism 90
Cokernel of a homomorphism 90
Combination, convex 107
  — absolutely 108
  — non-negative 109
Commuting operators 181
Compact set, Bohr 192
  — spectral 208
Complement of a subspace, algebraic 85
  — topological 157
Condition, boundary 11, 19, 71
  — distributed 71
  — initial 18
  — Levinson 209
  — Lipschitz 9, 163
  — of regularity of a functional 16
Cone 109
  — minihedral 118
  — total 186
Constant, interpolation 257
  — Lebesgue 55
Convergence, absolute 101, 135
  — mean square 13
  — unconditional 135
Convolution 22, 240
  — of sequences 27
Coordinates of a vector 86, 134
Corank of a homomorphism 90
Cross-property of a seminorm 173

Decomposition, triangular 220, 222
  — Wold 214
Defect of a homomorphism 90
Deficiency index of a symmetric operator 194
  — of an isometric operator 185
Definiteness, procedure for 230
δ-function 11
Derivative, Fréchet 163
  — generalised 244
  — of fractional order 42
  — Radon-Nikodým 142
Determinant, Fredholm 77
  — of a linear operator 170
Deviation 53, 147
278 Subject Index

Differential of a function 6
— of a mapping 163
Dilatation 226
Dimension, algebraic 86
— topological 136
Direct sum of linear operators 92
Disk algebra 234
Dissipation 18
Distribution 245
— of compact support 246
— of finite order 245
— probability 12
— regular 245
— tempered 250
Duality of linear spaces 147

Edge of a wedge 109
Equation, Abel integral 42
— characteristic 45
— Euler-Lagrange 16
— Fredholm integral 74
— heat 19
— Kepler 9
— Lagrange 17
— Laplace 243
— Poisson 247
— singular integral 82
— Volterra integral 74
— wave 18
Equimeasurability of functions 253
Equivalence of metrics 157
— of norms 158
— of operators 181
— of representations 180
— unitary 180
Extension of an operator 181

Flow 178
Formula, Cauchy-Hadamard 144
— Dirichlet 19
— Euler-Fourier 20
— Gauss 184
— interpolation
— Hermite 62
— Lagrange 62
— Newton 62
— inversion, Fourier 37
— Fourier-Plancherel 37
— Riemann-Mellin 39
— Weyl 49
— Widder 40
— Koch-Grothendieck 171
— Lagrange 72
— Lagrange-Sylvester 99

— of finite increments 163
— quadrature, Cotes 64
— Gaussian 59
— Taylor 7
Function, absolutely continuous 15
— absolutely monotonic 43
— almost-periodic 51
— analytic 7
— basic 245
— Cauchy 41
— characteristic 223
— differentiable 6, 163
— eigen- 51
— generalised 51
— flat 7
— generalised 245
— Green 71
— integrable 12
— interior 226
— Lagrange 16
— measurable 12
— normalised 13
— of compact support 38, 86
— periodic in mean 44
— positive 15, 36
— Riemann ζ- 42
— smooth 6
— summable 12
— locally 15
— Weierstrass 54
Functional, additive 11
— analytic 251
— bounded 127
— locally 118–119
— convex 110
— coordinate 88
— gauge 111
— Hermitian quadratic 77
— homogeneous 11
— positively 110
— linear 11
— Minkowski 111
— monotone 15
— multiplicative 104
— non-negative 238
— positive 15
— sesquilinear 196
— subadditive 109
— sublinear 110
— support 113, 115
— symmetric 238
— Hermitian 196
Functional calculus 99, 104, 170, 198, 208, 234
Subject Index

Gap between subspaces 137
Generalised sum of a series 100
Group, amenable 177–178
— dual 240
— one-parameter 186
— unimodular 179

Half-open interval 108
Half-space 113
Hamiltonian 46, 205
Harmonic oscillator 17
Homomorphism, absolutely summing 172
— bounded 153
— closed 155
— compact 165
— conjugate
— algebraic 90
— topological 155
— finite rank 93
— Hilbert-Schmidt 168
— Noetherian 93
— nuclear 169
— *- 238
Hull, convex 108
— absolutely 108, 138
— closed 138
— linear 88
Hyperfunction 251
Hyperplane 88
— of support 115
— separating 113
— almost 114
— strictly 128

Ideal, maximal 101
— standard 106
— simple 102
— symmetric 238
Index of a homomorphism 93
— of an indefinite metric 228
— of nilpotence 97
Inequality, interpolation 257
— triangle 13, 109
Infimum 118
Integral, Bochner 142
— Brodskij 221
— Dirichlet 243
— Fourier 36
— Geldand-Pettis 140
— Lebesgue 12
— Lebesgue-Stieltjes 15
— of Cauchy type 82
— of fractional order 42
Invariant mean 177

Jordan cell 97–98
— structure 97

Kernel, Cauchy 168
— difference 78
— Dirichlet 22
— Fejér 23
— Hermitian symmetric 77
— Hilbert-Schmidt 77
— of a homomorphism 89
— of a seminorm 111
— of an integral operator 71, 76
— Sushkevich 192

L-problem of moments 60
λ-measure 14
Lattice, Banach 234
— vector 118
— Dedekind complete 119
— Dedekind σ-complete 119
Logarithmic property of the index 93

Matrix, Hankel 30
— Heisenberg 206
— identity 86
— Jacobian 34
— Nevanlinna 33
— Toeplitz 35
Measure, Borel 14
— Choquet 143
— finite 12
— generalised 131
— Haar 179
— invariant 178
— Lebesgue 15
— Stieltjes 14
— of compact support 132
— operator-valued 198
— outer 14
— probability 12
— regular 116
— structure 242
— vector 142
— absolutely continuous 142
Method, Fourier 19
— of least squares 147
— of summation 100
— Cesáro 23
— Fejér 23
— of tangents 8
Metric, indefinite 227
— invariant 120
Model of a linear operator 222
— functional 223
Modulus of continuity 26, 254
Moment problem, abstract 116
   — Chebyshev-Markov 58
   — exponential 43
   — Nevanlinna-Pick 68
   — power 29
   — trigonometric 35
Monoid 178
Multiplicators 246
Multiplicity of an eigenvalue 199

Nil radical 102
Nilpotent 102
Nodes of a quadrature formula 59
Norm, graph 202
   — Hilbert-Schmidt 168
   — nuclear 169
   — of a function 13
   — of a homomorphism 127, 153
   — of a measure 132
   — of a vector 14, 111
   — of a vector measure 142
Normal eigenvalue 182
Normal mode 18
Number, approximation 168
   — defect 182
   — singular 216
Numerical range of a linear operator 215

Ω-topology 153
Operator, adjoint 192
   — algebraic 97
   — communication 227
   — compact relative to another operator 202
   — complete 216
   — completely non-selfadjoint 223
   — completely non-unitary 214
   — conservative 190
   — contracting 157
   — uniformly 157
   — damping 231
   — decomposable 210
   — differential 70, 259
   — elliptic 248
   — differentiation 11, 99
   — dissipative 190
   — dynamical 227
   — expanding 224
   — Hilbert 256
   — identity 11
   — infinitesimal 186
   — integral 71
   — Fourier 260
   — isometric 226
   — Laplace 18
   — linear 11, 91
   — metric 228
   — non-quasianalytic 209
   — normal 193
   — normally resolvable 162, 166
   — pseudodifferential 259
   — quasiconservative 212
   — quasidissipative 212
   — quasinilpotent 184
   — scattering 206
   — selfadjoint 73, 193
   — shift 99, 225
   — weighted 221
   — symmetric 46, 193
   — maximal 196
   — non-negative 196
   — unicellular 92, 221
   — Volterra 186
   — wave 205
Order of a pseudodifferential operator 259
   — of an eigenvalue 97
Parallelogram identity 130
Part of an operator 181
   — quasinilpotent 207
   — scalar 207
Pencil of operators, quadratic 231
   — strongly damped 231
Point, algebraic interior 108
   — extreme 138
   — interpolating 61
   — Lebesgue 23
   — quasiregular 182
   — regular 98, 181
   — stationary 10
   — strongly exposed 143
   — Weyl 32, 50
Polar 145
Polar representation of an operator 217
Polynomial, Bernstein 30
   — characteristic 41
   — Chebyshev
   — of first kind 54
   — of second kind 58
   — Faber 66
   — Hermite 28
   — Laguerre 28
   — Legendre 28
   — minimal 97
   — of best approximation 53
Principle, contraction mapping 9
   — fixed point 175
   — of duality 241
Subject Index

- of uniform boundedness 123, 154
- variational 10
Problem, Cauchy 164, 186
- - pointwise stable 188
- - stable 189
- Dirichlet 243
- of scattering theory
- - direct 204
- - inverse 204
- Riemann-Hilbert 82
- additive 82
- Sturm-Liouville 45
Process, interpolation
- - of Lagrange type 63
- - of Newton type 63
- Newton 8, 163
- quadrature
- - Cotes 64
- - Gaussian 64
Product, Blaschke 68
- of measures 15
- scalar 13
- tensor 173
Projection 91
- boundary 192
- root 98
Property, approximation 165
- of reciprocity of polars 148
- Radon-Nikodým 142
Pseudodistance 228
Pseudonorm 120
Quasimodel of a linear operator 222
Quasinilpotent 103
Quasinorm 121
Quasipolynomial 40
Quasiprojection 209
Radical of an algebra 102
Radius, spectral 184
- Weyl 31, 50
Random events 12
- variables 12
Rank of a homomorphism 90
- of a set of vectors 88
Ray 109
Regularisor 94
Representation 179
- almost-periodic 191
- bounded 179
- canonical functional 103
- contracting 179
- isometric 179, 233
- regular 180
- uniformly continuous 179
- unitary 179
Resolution
- of the identity 197
- spectral 98, 197, 199
Resolvent 103
- of a linear operator 98, 181
- - Fredholm 217
- of an element of an algebra 103
Restriction of an operator 181
Scale of Banach spaces 258
- normal 258
Schmidt expansion 217
Semiflow 178
Semigroup, amenable 177-178
- one-parameter 186
Seminorm 111
- factor 111
Separation of variables 19
Sequence, locally finite 245
- quasieigen- 182
- subadditive 184
- weight 25
Series, Dirichlet 42
- Fourier 20, 21, 52, 134, 136
- generalised 27, 45
- interpolation
- - Lagrange 68
- - Newton 62
- Neumann 162
- Taylor 7
Set, absorbent 110
- algebraically
- - open 108
- - solid 108
- balanced 108
- bounded 126
- Chebyshev 147
- circled 108
- conal 109
- convex 107
- - absolutely 108
- measurable 12
- precompact 154
- relatively compact 154
- - dense 52
- resolvent 98, 103, 181
- spectral 215
Shift, generalised 204
- semigroup 179
Shilov boundary 237
$\sigma$-algebra 12
Solution, fundamental 247
Subject Index

- generalised 248
- Banach 14
- barrelled 122
- configuration 16
- conjugate
- - algebraic 88
- - topological 129
- coordinate 134
- countably Hilbert 173
- defect 182
- factor 88
- Fréchet 122
- fully complete 152
- function 11
- Hilbert 14
- intermediate 256
- interpolation 257
- linear topological 119
- - complete 122
- locally bounded 126
- - convex 124
- Lorentz 253
- Mackey 149
-Marcinkiewicz 253
- maximal ideal 103
- Montal 150
- Nikolskij-Besov 254
- normed 14
- nuclear 171
- of Lipschitz type 254
- ordered 117
- Orlicz 253
- phase 178
- pre-Hilbert 14
- probability 12
- Pták 152
- reflexive algebraically 90
- - topologically 130, 150
- Riesz 118
- Schwartz 250
- second conjugate 150
- Slobodetskij 254
- smooth 130
- Sobolev 254
- strictly convex 131
- super-reflexive 158
- symmetric 257
- uniformly convex 130
- smooth 130
- Spectrum 51, 97, 103, 181
- - absolutely continuous 200
- approximate 182
- Bohr 52
- combined 236

- continuous 51, 199
- discrete 51, 199
- residual 182
- simple 224
- singular 199-200
Subordination of metrics 158
- of norms 158
Sum of spaces, external direct 89
- - orthogonal 87
Sum of subspaces, direct 87
- - - topological 157
- - - orthogonal 157
Support of a distribution 246
- of a function 86
- of a measure 143
- of a vector 85
- of an analytic functional 251
Supremum 118
Symbol of a differential operator 248, 259
- of a pseudodifferential operator 259
System, biorthogonal 61
- Chebyshev 56
- complete 27, 132-133
- distal 136
- dynamical 178
- Markov 57
- - normalised 59
- minimal 135
- - linearly independent 134
- orthogonal 13
- orthonormal 13, 134
Theorem, boundary spectrum splitting-off 192
- closed graph 155
- convolution 22
- ergodic 191
- expansion 40
- implicit function 9
- inverse homomorphism 156
- local 235
- open mapping 156
- separating hyperplane 113, 114, 128
- spectral 98, 197, 200, 230
- spectral mapping 99, 104, 207, 209, 235
- stability of the index 95
- superposition 209
Theory, Fredholm 76
- Hilbert-Schmidt 77
- Noether 84
- scattering, non-stationary 205
- stationary 204
Topological isomorphism 121
Topology, consistent with duality 148
<table>
<thead>
<tr>
<th>Subject Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>— Dieudonné-Schwartz 245</td>
</tr>
<tr>
<td>— locally convex 124</td>
</tr>
<tr>
<td>— Mackey 149</td>
</tr>
<tr>
<td>— norm 150</td>
</tr>
<tr>
<td>— of convergence in measure 120</td>
</tr>
<tr>
<td>— of uniform convergence 125</td>
</tr>
<tr>
<td>— on bounded sets 150</td>
</tr>
<tr>
<td>— on compact sets 125</td>
</tr>
<tr>
<td>— strong 145</td>
</tr>
<tr>
<td>— tensor product 173</td>
</tr>
<tr>
<td>— uniform 150, 154</td>
</tr>
<tr>
<td>— weak 145</td>
</tr>
<tr>
<td>— weakened 145</td>
</tr>
<tr>
<td>Total family of linear functionals 124</td>
</tr>
<tr>
<td>— of projections 92</td>
</tr>
<tr>
<td>— of seminorms 124</td>
</tr>
<tr>
<td>Trace 170</td>
</tr>
<tr>
<td>— matrix 170</td>
</tr>
<tr>
<td>— spectral 219</td>
</tr>
<tr>
<td>— tensor 169</td>
</tr>
<tr>
<td>Transform, Cayley 194</td>
</tr>
<tr>
<td>— Fourier 36</td>
</tr>
<tr>
<td>— generalised 49, 248</td>
</tr>
<tr>
<td>— operator 210</td>
</tr>
<tr>
<td>— -Plancherel 37</td>
</tr>
<tr>
<td>— -Stieltjes 37</td>
</tr>
<tr>
<td>— Geland 233</td>
</tr>
<tr>
<td>— Laplace 39</td>
</tr>
<tr>
<td>— finite 44</td>
</tr>
<tr>
<td>— local 43</td>
</tr>
</tbody>
</table>

| — — — -Stieltjes 42 |
| — Legendre 252 |
| — Travelling wave 18 |
| Ultradistribution 250 |
| Ultrapower 106 |
| Unit sphere 127 |
| Variation of a functional 16 |
| Vector, channel 223 |
| — cyclic 224 |
| — root 97 |
| Vector-function, absolutely integrable 141 |
| — analytic 144 |
| — Bochner integrable 142 |
| — entire 144 |
| — Geland-Pettis integrable 140 |
| — simple 141–142 |
| w*-topology 144 |
| Wedge 109 |
| — generating 109 |
| Weight 28 |
| Weights of a quadrature formula 59 |
| Weyl disk 32, 50 |
| — point 32, 50 |
| — radius 31, 50 |
| Young relation 132 |